12-2019

Gesture-based profiling of commonplace lifestyle and physical activity behaviors

Meeralakshmi RADHAKRISHNAN
Singapore Management University, meeralakshm.2014@phdis.smu.edu.sg

GESTURE-BASED PROFILING OF COMMONPLACE LIFESTYLE AND PHYSICAL ACTIVITY BEHAVIORS

MEERALAKSHMI RADHAKRISHNAN

SINGAPORE MANAGEMENT UNIVERSITY

2019

Gesture-based Profiling of Commonplace Lifestyle and Physical Activity Behaviors

by

Meeralakshmi Radhakrishnan

Submitted to School of Information Systems in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Computer Science

Dissertation Committee:

Archan MISRA (Supervisor / Chair)
Professor of Information Systems
Singapore Management University

Kotaro HARA
Assistant Professor of Information Systems
Singapore Management University

Hwee Xian TAN
Assistant Professor of Information Systems (Practice)
Singapore Management University

Inseok HWANG
Research Staff Member, Master Inventor
IBM

Singapore Management University
2019

I hereby declare that this PhD dissertation is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in this dissertation.

This PhD dissertation has also not been submitted for any degree in any university previously.

Meeralakshmi Radhakrishnan

9th December 2019

Gesture-based Profiling of Commonplace Lifestyle and Physical Activity Behaviors

Meeralakshmi Radhakrishnan

Abstract

The widespread availability of sensors on personal devices (e.g., smartphones, smartwatches) and other cheap, commoditized IoT devices in the environment has opened up the opportunity for developing applications that capture and enhance various lifestyle-driven daily activities of individuals. Moreover, there is a growing trend of leveraging ubiquitous computing technologies to improve physical health and well-being. Several of the lifestyle monitoring applications rely primarily on the capability of recognizing contextually relevant human movements, actions and gestures. As such, gesture recognition techniques, and gesture-based analytics have emerged as a fundamental component for realizing personalized lifestyle applications.

This thesis explores how such wealth of data sensed from ubiquitously available devices can be utilized for inferring fine-grained gestures. Subsequently, it explores how gestures can be used to profile user behavior during daily activities and outlines mechanisms to tackle various real-world challenges. With two daily activities (*shopping* and *exercising*) as examples, it then demonstrates that unobtrusive, accurate and robust monitoring of various aspects of these activities is indeed possible with minimal overhead. Such monitoring can then, in future, enable useful applications (e.g., *smart reminder* in a retail store or *digital personal coach* in a gym).

First, this thesis presents the *IRIS* platform, which explores how appropriate mining of sensors available in personal devices such as a smartphone and a smartwatch can be used to infer micro-gestural activities, and how such activities help reveal latent behavioral attributes of individual consumers inside a retail store. It first investigates how inertial sensor data (e.g., accelerometer, gyroscope) from a smartphone can be used to appropriately decompose an entire store visit into a series of modular and hierarchical individual interactions, modeled as a sequence of

in-aisle interactions, interspersed with non-aisle movement. Further, by combining such sensor data from a wrist-worn smartwatch and by deriving discriminative features, the *IRIS* platform demonstrates that different facets of a shopper's interaction with individual items (e.g., picking an item, putting an item in trolley), as well as attributes of the overall shopping episode or the store, can be inferred.

This thesis next investigates the possibility of using a *wearable-free* sensing modality for fine-grained and unobtrusive monitoring of multiple aspects of individuals' gym exercises. It describes the *W8-Scope* approach that requires no on-body instrumentation and leverages only simple accelerometer and magnetometer sensors (on a cheap IoT device) attached to the weight stack of an exercise machine to infer various exercise gestures, and thereby identify related novel attributes such as the amount of weight lifted, the correctness of exercise execution and identify the user who is performing the exercise. It then also experimentally demonstrates the feasibility of evolving W8-Scope's machine learning-based classifiers to accommodate the *medium-time* scale (e.g., across weeks or months) changes in an individual's exercise behavior (an issue that has received insufficient attention to date).

Finally, this thesis explores the possibility of accurately inferring complex activities and gestures performed concurrently by *multiple individuals* in an indoor gym environment. It introduces a system that utilizes a hybrid architecture, combining sensor data from *'earables'* with non-personal IoT sensors attached to gym equipment, for individual-specific fine-grained monitoring of weight-based exercises in a gym. Using real-world studies conducted with multiple concurrent gym-goers, this thesis validates that accurate association of "user-equipment" pairings is indeed possible, for a majority of common exercises, in spite of the significant signal dampening on the earable. Moreover, it demonstrates how features from the earable and IoT sensors can be combined to significantly increase the accuracy and robustness of exercise recognition. In future, the real-time exercise analytics capabilities developed in this thesis can be used to enable targeted and individualized real-time feedback on user dynamics and increase user engagement.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I'm indebted to several individuals who have been phenomenal in helping me sail through this long and arduous journey and successfully complete this effort.

First and foremost, I would like to express my heartfelt gratitude to my advisor, Professor Archan Misra, for his guidance, kindness and tremendous support throughout all these years. I consider myself extremely fortunate and blessed to have been advised by you and cannot ever imagine having a better mentor than you. It has not been an easy journey and words cannot do justice in explaining how thankful I'm to you for being so understanding, caring and supportive. Your dedication, passion for work, conscientiousness, compassion and thoughtfulness towards everyone, has certainly been inspirational. Thank you for being patient with me, having shown faith in me and for always being encouraging throughout this marathon endeavor. I couldn't have completed this Ph.D. without your enduring guidance and I'm deeply grateful to you for everything that you've done for me and is continuing to do for me. Whatever I've achieved, whomever I've become and will become professionally, I owe it to you. You are indeed exceptional and the best!

I'm also truly grateful to my wonderful dissertation committee. I thank Kotaro Hara, Hwee-Xian Tan and Inseok Hwang for kindly agreeing to serve on my dissertation committee. It has indeed been a humbling experience and I thank each of you for providing me insightful comments and extremely valuable feedback to improve the quality of my dissertation. Inseok, I treasure your commitment to research, your kind words and the support you've extended to me. My sincere thanks to Rajesh for helping me in my initial years of research and I've learned a lot about systems research from you. I'd like to extend my deepest thanks to Youngki Lee, Sharanya

Eswaran, Asim Smailagic, Daniel Siewiorek, Koustuv Dasgupta, Deepthi Chander for their guidance at various phases of this journey. Sharanya, you're truly inspirational, I feel fortunate to have been mentored by you during my internship in the first year and I cherish all the good times we've had together.

I'm indebted to the generous support extended by SMU's PhD Full Scholarship, LiveLabs and LARC which helped me to expand my horizon and travel the world. I'm thankful to Baihua, Pei Huan, Caroline, Yong Fong, Sipei, Kazae, Francis, Barbara, Phoebe and Lydia for all their incredible support.

I feel privileged and happy to have shared my journey at SMU with amazing friends from the lab. Kasthuri, Jeena, Sougata, Swetha, Madhumita, Kartik, Amit, Vignesh, Thivya, Swapna, Camellia, Chau, Vu, Sinh, Loc, Huy, Ratan, Dulanga, Lakmal, and Tuan, I'll forever cherish our times together, the memories of gaming evenings, b'day celebrations and a lot more. Kasthuri, a very special mention to you, but I'm short of words to express how grateful I'm to have you as my best buddy and forever go to person from day one at SMU till now. Thank you and a big hug for being there for me through thick and thin. Jeena, thank you for everything and for considering me like your own family. Sougata, you've helped me a lot, thank you for being there for me always like a brother. I'm also thankful to my friend Athul for consistently encouraging and cheering me on throughout this journey.

Finally, I sincerely thank my family for their love and support. I'm grateful to Amma and Achan for their unconditional love. You both have shown me and Chettan how to persevere through all kinds of tough situations. Chetta, all of these happened only because you motivated and instilled in me this idea of doing a Ph.D. in Singapore. You've always been my source of inspiration and I'm happy to have followed you in stepping out of the comfort zone of home to a new country to pursue my passion. Malu, thank you for encouraging me and reassuring to me that things would fall in place with time. This journey would not have been possible without all your patience, constant encouragement and support.

To

Amma, Achan and Chettan

for a lifetime of unconditional love and support

Chapter 1

Introduction

1.1 Improving Daily Lifestyle through Unobtrusive Technologies

In recent years, there has been an increasing research interest in utilizing sensor data from personal devices (smartphones and wearables) and Internet-of-Thing (IoT) sensors to automatically recognize mainstream daily lifestyle activities (e.g., *exercising* [25, 77, 89, 141], *eating* [126, 108, 73] or *sleeping* [27, 72, 81]). Notably, the global consumer trends continue to indicate the popularity and growing number of personal smart devices carried by individuals in their daily life. Reports from Statista [5] and CCS Insight [7] forecast that worldwide smart wearable device sales will double by 2022 (with 85 million smartwatches to be sold in 2019 and that increasing to 137 million in 2022). Digitally capturing and automatically sensing different aspects of human daily lifestyle yield substantial opportunities in leading a better quality of living (for example, automatically capturing an individual's workout activity may help them to maximize their workout effectiveness) by reducing human effort, saving time and capturing information otherwise not readily trackable. In the years ahead, the combination of machine learning, artificial intelligence (AI) and contextually rich data streams delivered by personal devices and IoT sensors would provide extensive opportunities to enable exciting real-life applications that make our lives smarter and easier.

In this thesis, I describe a set of novel techniques and analytics pipeline for inferring fine-grained human actions and gestures and applying machine learning on such inferences to derive more quantified insights on individuals' lifestyle activities. The proposed techniques leverage multiple sensors available in (a) individual's personal devices (e.g., smartwatch, earphones) or (b) cheap IoT devices (e.g., inertial sensor units) deployed in the environment or that can be attached to common objects, or (c) a combination of those two types of devices. In addition to enabling fine-grained and accurate monitoring of lifestyle activities, I also focus on making these systems and services unobtrusive, cheap, easy to deploy, convenient to use.

The ability to identify a person's movements, determine what gestures they may be performing and profile their overall behavior is especially useful in the wellness and retail domains. As such, in this thesis, I focus on exploring the possibilities of recognizing gestures and monitoring other fine-grained aspects of two everyday lifestyle activities: *shopping* and *exercising*. This thesis builds upon several of the existing mobile/wearable/IoT-based gesture recognition techniques previously (or contemporaneously) proposed in literature, and demonstrates the judicious use of a combination of sensors for accurate monitoring of aspects of these two lifestyle activities, either at finer granularity than previously possible or including attributes that have previously not been monitored. I also determine the extent to which personal devices are sufficient for monitoring certain activities, and the additional advantages and convenience that can arise from the use of simple devices/sensors that are already deployed in the indoor space for other purposes.

1.2 Motivating Scenarios

From Section 1.1, it is evident that unobtrusive technologies that monitor daily lifestyle activities will be useful to individuals. To strengthen the case further, let us consider these motivating scenarios in different contexts. These scenarios are derived based on formative studies (i.e., observational studies performed at retail grocery stories and gyms) conducted at different stages of the research process.

1.2.1 Scenario 1: In-store Retail Insights on Shoppers

Joe runs a grocery store and is planning to reform it as a *smart retail store* to enable smart retail applications, streamline the customer experience and thus, improve his business (note that this does not involve business process improvements (e.g., in terms of logistics, operations, etc.)). He is fascinated by the idea of capturing shopping actions and browsing behaviors of consumers inside the store, as such behavior will allow him to improve the customer's shopping experience–for example, by making necessary changes to the store layout or item arrangements, or by providing proactive recommendations to shoppers who are in a rush. However, Joe does not wish to make any changes to the store's IT infrastructure (at least initially) and would like to realize his vision with minimal expenditure. He consults an IT firm with expertise in providing sensor-based solutions for retail businesses. As individuals are increasingly carrying personal devices, such as smartphones and smartwatches, during their daily lifestyle activities, the IT firm advises Joe of the high potential in utilizing the sensors embedded in these personal devices to infer individual level consumer behavior. With the help from the IT firm, Joe develops a custom *smart retail* application (which records sensor data from shopper's personal devices and also has the capability for shoppers to include their shopping list and item preferences) for his store. Joe incentivizes the customers with a 5% discount on their bill to download and use this application.

To realize this scenario, an in-store shopping behavior monitoring system should be able to perform the following:

- capture sensor data from customer's smartphone and smartwatch using the *smart retail* application,

- construct generalized shopping gesture recognition models based on the customer's shopping data recorded,
- utilize these models to first determine the movement behavior of the shopper inside the store (i.e., whether the shopper is in a non-aisle or aisle zone),
- identify shopper's gestural interactions with items of interest (such as picking up an item, returning an item back to the shelf, putting an item into the trolley),
- use these item-level observations (i.e., how many items the shopper placed in her cart, time spend browsing an item etc.), to understand a shopper's overall behavior (for example, whether the shopper is in a hurry, if the shopper is picking a familiar item etc.),
- obtain longitudinal insights, across multiple store visits, from individual customers to build a shopper profile for targeted advertising and smart reminders.

In envisioning this scenario, we believe that our proposed approach of using data sensed from 'only' personal devices of the shopper is an appropriate option to enable fine-grained, cost-effective and personalized insights based on an individual's shopping behavior. We also believe that our approach has the advantage of identifying specific item-level interactions compared to other emerging/state-of-the-art alternatives (e.g., Wi-Fi APs, BLE beacons) which only provides shopper's location and interest. Moreover in a scenario like 'shopping', individuals are more likely to be carrying their personal devices and also as most of the significant shopping actions are performed by the arms, a wrist-worn smartwatch would be able to capture these behaviors in a holistic manner (as opposed to during activities like 'exercising' where there is a lot of both upper and lower-body actions are involved and then a wrist-worn device may not be able to capture the lower-body movements). In addition, this paradigm requires no significant IT capital or operational investments by the store operator, and might represent one possible approach for capturing shopper insights in smaller or regional retail stores in expenditure conscious emerging economies (e.g., in Africa or Asia). Additionally, video analytics using in-store cameras captures behavior of collective buyers but not individuals or requires extensive and expensive instrumentation (e.g., Amazon Go [1]). In the future, as Internet-of-Thing (IoT) deployments become more widespread, our proposed technique can be potentially integrated with solutions such as Amazon Go to enable additional services such as providing real-time promotions or information prompts based on the exact item a shopper picked or interacted, checkout-free shopping experiences.

1.2.2 Scenario 2: Quantified Insights on Weight Stack-based Gym Exercises

Annie has been going to the gym for the past few months and has recently started training on an exercise machine with stacked weights. She wants to reduce the risk of any injury and maximize the effectiveness of her workout by tracking her daily exercises in an easy and convenient way. However, she is not able to afford a personal trainer neither does she likes to wear any device on her body while exercising. She learns from her friend that the weight machines in the gym are now equipped with some simple sensors, which enables the gym to automatically track an individual's exercises and then provide quantified and personalized insights (*like a personalized digital coach*). This technology excites Annie, who decides to sign up and use the personalized Web portal to track her machine exercises and get personalized feedback. After a few day, Annie is able to see her exercise summary and the details of her daily workout such as the amount of weight lifted, number of exercise repetitions/sets for all the exercises she performed on weight machines.

To realize this scenario, the machine exercises monitoring system should be able to perform the following:

- record and centrally store the sensor data when Annie is performing different exercises on the weight stack-based machine,

- extract features and construct personalized models based on Annie's workout data,
- when the next time Annie uses the exercise machine, automatically identify that it is Annie who is currently using the exercise machine,
- infer various insights from her workout such as the time taken to complete a repetition, the exercise performed (e.g., triceps pushdown), the weight lifted (e.g., 8kg),
- provide quantified exercise summary and personalized feedback to Annie, based on the inferred insights, at the end of her workout,
- continue to track Annie's activities over longer time scales, even though Annie's exercise patterns may change as a result of her evolving physique or exercise familiarity (e.g., on a certain day if she performs an exercises too fast or lifts a much heavier weight than her usual) and evolve the models to adapt to these evolutionary changes in behavior.

In contrast to Scenario 1, in the above scenario of monitoring weight-based machine exercises in a gym, we propose a wearable-free and unobtrusive approach of leveraging only cheap IoT sensors attached to the exercising equipment. The choice of this specific sensing modality is based on the following key reasons: (a) reluctance of individuals to adopt on-body devices during exercise activity (e.g., due to the discomfort in strapping devices to the body (mainly as people sweat while exercising) and also because it is inconvenient to carry devices like smartphone on the body while performing exercises), (b) limitations of wearable devices in capturing both upper and lower-body exercises (e.g., a wrist-worn smartwatch would not be able to capture patterns of leg exercises) and (c) potential privacy concerns arising from the use of video sensing techniques. Thus, the main advantage of our proposed approach is that without requiring any instrumentation on the individual's body and without using video-sensing, it can still obtain detailed insights on the exercises performed on weight machines. While in our approach we intend to use simple inertial sensors (e.g., accelerometer and magnetometer on an off-the-shelf IoT device) to capture the exercising aspects, there are also other potential sensors that could be used. For example, laser gauges or distance sensors (attached to the top of the weight stack of exercise machine) could be used to measure the distance moved by the weight stack as well as the amount of weight being lifted (based on the change in placement of the pin attachment to the respective weight slab).

1.2.3 Scenario 3: Multi-user Exercises Monitoring in a Gym

Roy is the head of the technology division of the Sports and Wellness Center. This center focuses on empowering people to live a healthy and active lifestyle and provides technological solutions to help them. Roy along with his team is currently in the process of transforming the public gyms to be smart gyms, that provide a more personalized and engaging experience to the gym users. More specifically, they want to roll out some technology at all the gyms as a way to capture exercises performed simultaneously by multiple users, identify what each individual is doing, provide them quantified insights, personalized feedback and corrective actions. However, he only has a limited budget to make it happen and therefore, procuring sophisticated exercise machines with in-built sensor technology is not a scalable option. In addition, gym-goers often utilize smaller equipment, such as dumbbells, barbells. To extend the quantified analysis of exercises to the potentially hundreds of such small-form factor gym equipment, it is important to employ only inexpensive sensors. Two other plausible options in front of him are to (i) utilize the video feeds from surveillance cameras in the gym or (ii) rely on custom-wearable devices attached to specific parts of the body (e.g., the limbs, chest) with the goal of capturing individual exercise behavior. But based on his interactions with users of the gym as well as from discussion with other technology experts, Roy realizes that each of these technologies has several limitations and special adoption challenges. For example, deploying video sensors in gym areas are likely to raise privacy concerns; similarly, users are likely to be reluctant to strap on on-body sensing devices while exercising, unless of course the wearable is already used for other commonplace lifestyle reasons (e.g., earphones for listening to music). To adhere to such privacy and budgetary constraints, Roy thus has to devise solutions that exclude the use of video cameras, but can leverage on the attachment of small form-factor sensors attached to individual exercise equipment, such as dumbbells. In the process of brainstorming various other technologies, Roy came across the recent advances in "earable" technology, where sensors embedded in ear-worn devices can be used to monitor different human activities. He is excited by the fact that earphones are commonly used by exercising individuals and it also can enable personalized and real-time audio feedback. The team then decides to combine this technology with their already existing solution of utilizing small form-factor sensors attached to the exercise equipment to distinguish between multiple people exercising in the gym and capture fine-grained aspects of each user's exercising behavior.

To realize this scenario, a smart gym multi-user exercises monitoring system should be able to perform the following:

- an application to capture sensor data from gym equipment-attached sensors and personal sensor-enabled earphones of multiple individuals exercising in the gym,
- establish the *association* between an individual's earable device and the corresponding gym equipment–i.e., pair a user with the impersonal gym object which he or she is currently interacting,
- recognize the exercise gestures/movements and identify the type of exercise performed by each individual using this pair of (earable, equipment) sensor data,

- create personalized profile for individual gym users and map their exercise-specific insights and display their exercise data, summary reports etc. using a smartphone application or a web portal.

- generate personalized real-time audio-based feedback based on each individual's exercising behavior (e.g., alerting the user to slow down if he is going too fast)

In the above scenario, leveraging a combination of off-the-shelf IoT sensors plus wearable sensors seem to be an appropriate option to realize a cost effective and a simple system to enable real-time personalized feedback to individuals. As briefly discussed earlier, our choice of ear-worn sensors is motivated by the key fact that they are a more socially acceptable class of wearable devices and people commonly use earphones while working out in a gym. They are also small in form-factor, unobtrusive and may not cause discomfort (e.g., from sweating or from restriction in freely doing the exercises) unlike other kinds of wearable devices (e.g., a smartwatch or a chest strap) strapped to other parts of the body.

1.3 Pervasive Sensing for Gesture Recognition in Lifestyle Monitoring Applications

As described in the above motivating scenarios, automated and unobtrusive recognition of various gestures/actions performed as part of daily activities can be beneficial to the individuals (to the end users as well as business owners) in many ways. Moreover, the availability of multitude of sensors in different devices that are readily available with the individuals or available in the environment opens up unprecedented opportunities in realizing several such useful applications. Building upon previous works that utilizes sensor-based techniques for daily activity monitoring, this thesis demonstrates the use of innovative sensing modalities and novel machine-learning based analytics pipelines that specifically target these two key application domains.

1.3.1 Leveraging sensors on personal devices

The availability of multitude of sensors on our personal devices such as smartphones, smartwatches, earphones, now offers the possibility of capturing rich and varied information of human context. In other words, these personal devices have become extensions of ourselves (i.e., "what, where, when and why" people do certain activities). In Section 2.1, I explore the potential of using a combination of smartphone and smartwatch sensors (e.g., accelerometer, gyroscope, step counter) to first infer a shopper's in-store micro-gestural activities, such as "picking up an item" or "placing it in a shopping cart". Then, I use the observed pattern of such gestures to infer a shopper's higher-level profile, such as "the shopper is in a hurry" or "shopper is familiar with the store". We believe that our proposed solutions, for capturing latent in-store individual behavior are practical and attractive as they can work *without* requiring infrastructure support, such as Wi-Fi APs, BLE beacons or in-store cameras.

1.3.2 Leveraging sensors on IoT devices in the environment

Even though it is likely that our personal devices are with us during most part of our daily life, there are certain contexts or situations when our smartphones or smartwatches become inconvenient to carry or are unable to provide comprehensive observability (e.g., while exercising in a gym). In such cases, tapping the ubiquitously available, cheap and simple Internet-of-Things (IoT) based sensor devices present significant opportunities. These devices can be attached to objects that individuals interact with or can be deployed in the environment. In Section 3.2 and Section 4.1, I demonstrate the use of magnetometer and accelerometer sensor-equipped cheap IoT devices that can be attached to either exercise machines or dumbbells to obtain fine-grained aspects of an individuals' weight-based gym exercises.

1.3.3 Fusing sensors from personal and IoT devices

In certain scenarios, simply using data sensed from either the individual's personal device or the IoT devices in the environment may not necessarily be sufficient, especially (i) when there are multiple individuals in the environment, (ii) when individuals do not interact with a very limited set of discrete objects, but interchangeably use many objects (e.g., dumbbells), and (iii) when the applications need to capture finer-grained aspects of each individual's activity, with minimal intrusion, and also execute real-time interventions. As such, in Section 4.1, I demonstrate the feasibility of using data sensed from unobtrusive wearables (such as earphones), combined with sensor data from devices (e.g., inertial measurement units) attached to the exercise equipment, to identify exercising aspects of multiple people in the gym and provide them with real-time personalized corrective feedback.

1.4 Key Challenges

In order to realize the aforementioned real-life scenarios of monitoring various activities of daily lifestyle by fusing data from multiple sensors, numerous challenges have to be addressed. I list down some of those challenges below:

1. **Accuracy:** Accurately recognizing various human actions and gestures performed as part of different activities is vital for ensuring the practical acceptance of any end-user application. In real-life scenarios, meeting high accuracy requirements can be challenging and is highly dependent on the appropriate choice of sensors used, amount of training data available and the classification models used. For example in Scenario 1 (described earlier), the in-store shopping activity recognition system should accurately segment the movements and different item-level interactions and gestures of the shopper

to obtain an overall understanding of the shopping behavior. In this case, judiciously mining the sensor data from both smartphone and smartwatch of the shopper is important to obtain the required insights accurately. Similarly, it is important for an exercise recognition application to accurately identify the different exercises and related aspects. Novel analytics pipelines and sensor fusion techniques are required to obtain accurate inferences.

2. **Sensing at Finer-granularity:** The usefulness of several of the lifestyle monitoring applications relies also on the granularity of the information that can be sensed and the variety of insights that can be provided to the individual. For example in Scenario 2, the exercise monitoring system needs to identify each of the exercises and the intensity at which Annie performs exercises in a gym session in order to provide her a comprehensive exercise summary, as well as personalized recommendations and feedback. Additionally, such a system should also be able to monitor exercise-mistakes, both at set-level and also within a set (e.g., for enabling real-time feedback through a personal earable device).

3. **Robustness to Time Varying Changes in User Behavior:** One of the key aspect of most of the activity monitoring systems is its dependency on user behavior. However, individual behaviors/styles are prone to changes and most of the lifestyle monitoring systems are built with training data collected over relatively short observational periods. In applications such as exercise behavior monitoring, it is important for the system to be robust enough to capture the inherent within-user differences (i.e., adapting the models to medium time-scale changes in individual behavior).

4. **Sensor Location:** Another key challenge associated with practical application scenarios is the appropriate location and placement of sensor devices. The sensor devices (either attached to human body or available in the environment) can be exposed to noise, interference, and other confounding effects

caused by nearby objects and users. Additionally, certain applications may require placement of sensors at locations where it is difficult to comprehensively capture all required motion patterns. For example, with the unfavorable on-body placement of earables, it is indeed questionable whether ear-based inertial signals can provide any discriminative information about exercise motion, especially when such motion is primarily restricted to upper or lower limbs. Similarly, data obtained from sensors on the top of a weight stack may be noisy (e.g, interference on the magnetic sensor from the dumbbells carried by nearby users) and affect the system's performance.

5. **Discriminating Accurately in a Multi-user Environment:** In scenarios where there are multiple people performing various gestures/activities in the environment and also when using individualized wearable devices are not readily feasible, it is important to discriminate between individuals to perform personalized monitoring. For example, in a gym environment where multiple people are performing different exercises without necessarily wearing multiple on-body devices, we will need to develop solutions that can unobtrusively distinguish among multiple individuals.

6. **Energy Consumption:** Minimizing the energy consumed, especially by personal devices (such as smartphones and smartwatches) while performing the required sensing to provide accurate and fine-grained activity monitoring is one of the key challenges in pervasive applications. As these devices have limited battery capacity and are not merely intended to just perform these analytics, it is important to save their energy to perform other primary tasks. Some common approaches to minimize the energy overhead is by incorporating mechanisms such as duty cycling the sensing operation or adaptive sensing based on certain inferences.

7. **Privacy:** Although the proposed techniques and solutions in this thesis aims to minimize the associated privacy concerns (compared to systems that rely

on video/audio sensing for such fine-grained monitoring), collecting sensor data, especially, from personal devices naturally raises some level of privacy concerns. Therefore, it is important to consider the privacy aspects and have appropriate mechanisms in place to tackle common privacy threats.

This thesis aims to address the first five challenges outlined above.

1.5 Motivating Human Activity Recognition (HAR) Research

Due to its immense potential in providing personalized support for many different applications and fields of study (such as medicine, sociology, human-computer interaction, or human security), Human Activity Recognition (HAR) continues to be an area of active research. Researchers have explored different modalities and proposed techniques that are mobile/wearable sensor-based [66, 135, 27, 33, 84], vision-based [93, 63, 140, 108], or wireless sensing-based [131, 138, 100, 128] to obtain varying levels of insights on specific activities of people. Well known examples include the RiSQ system [84], for identifying smoking gestures using smartwatch sensors, ThirdEye [99] for tracking browsing behaviors of shoppers using a smartglass and WiFi, WiSee [96] for whole-home gesture recognition, WiBreathe [100], a wireless system for estimating human respiration rates etc.

Although several such techniques have been proposed for HAR, there are still open opportunities and limitations that are not addressed. Existing approaches that are targeted at detecting specific activities still face several challenges: (a) attaining high performance accuracy even while sensing activities at finer granularity, (b) being unobtrusive, simple and cheap, (c) robustness in working in real-world conditions and over longer time scales and (d) ability to accurately recognize activities of multiple individuals in the environment. In this dissertation, I focus on monitoring two key human activities: *shopping* and *exercising* and introduce solutions that tackle several of these aforementioned challenges not dealt with in prior works in the similar domain.

1.6 Thesis Statement

Previous sections highlight the opportunities that arise from the availability of different sensors on personal and other IoT devices and some of the key challenges involved in enabling different lifestyle monitoring applications put forward. In this dissertation:

> I demonstrate that it is feasible to combine novel machine learning-based analytics techniques with judicious fusion of sensor data from commodity mobile, wearable and/or IoT devices to: (a) accurately recognize human shopping and exercise-related gestures, under real-world diversity and usage artefacts, at finer granularity, (b) use such gestural inferences as building blocks to derive useful and robust higher-level insights about an individual's shopping and exercising behavior and, (c) ensure that such machine learning-based activity inferencing models perform robustly in the face of medium-term evolution in an individual's behavior.

This dissertation establishes the thesis through the following steps:

1. First, it presents the opportunities in exploiting the richness of human actions and gestures involved while performing commonplace daily lifestyle activities. Using two everyday lifestyle activities: *shopping* and *exercising* as examples, it identifies the complexities and characteristics that are unique to each activity and determines the design goals and the challenges involved in realizing such lifestyle monitoring applications at finer granularity.

2. To demonstrate the applicability of using data sensed from multiple personal devices for fine-grained activity monitoring and user profiling, it then presents

the *IRIS* platform that uses standard locomotive and gestural micro-activities as building blocks to define novel composite features that help classify different facets of a shopper's interaction/experience with individual items, as well as attributes of the overall shopping episode or the store.

3. It then presents the *W8-Scope* system that utilizes only a simple, cost-effective sensor, containing only a 3-axis accelerometer and a 3-axis magnetometer, mounted on the weight stack of gym exercise machines, to obtain fine-grained insights into multiple aspects of individual's gym exercise behavior. To motivate this application and the chosen approach, it presents results of analysis of both digital gym usage records as well as a survey of 575 gym-goers. Moreover, it also demonstrates that by adopting incremental learning techniques, *W8-Scope* can accurately track various facets of exercises over longitudinal periods, in spite of the inherent within-user differences that occur in exercising behaviors.

4. Finally, this dissertation explores the possibility of simultaneously extracting gestural insight of multiple active users, based on a combination of wearable and IoT sensors. It uses free-weights exercises monitoring of multiple users in a gym as an example scenario. In particular, it develops novel techniques that combine inertial data sensed from a common personal lifestyle device (e.g., earphones), and IoT devices attached to the exercise equipment to distinguish between multiple individuals and infer his/her exercising behavior.

Chapter 2

In-Store Shopper Behavior

In this chapter, I demonstrate the capability of leveraging data fused from multiple sensors in individual's personal devices such as a smartphone and a smartwatch to obtain a detailed understanding of an individual's in-store shopping behavior. With real world studies conducted at two retail stores, I validate the proposed approaches in accurately capturing various item-level gestural interactions of shoppers and how such inferences can be used to derive further insights on shopper's behavior.

2.1 Capturing In-Store Retail Insights on Shoppers

Faced with increasing online competition, retail store owners are increasingly interested in the ability to better understand the browsing behaviors and intentions of consumers inside their physical stores. A variety of technologies, such as Wi-Fi and BLE beacon-based aisle-level location tracking [118], RFID based asset monitoring [109] and smartglass-based browsing monitoring [99] have been explored to capture such individual and collective in-store behavior. While these advanced technologies hold great promise, their cost makes them unlikely to be adopted widely, especially in low-margin, emerging economy markets (such as India, China or Brazil) in the near future.

We believe that solutions for capturing latent in-store individual behavior become much more practical if they can work without requiring infrastructure support, such as Wi-Fi APs, BLE beacons or in-store cameras. Moreover, as discussed previously, there is a great potential in tapping the sensors available on personal smart devices for monitoring individual behavior. Accordingly, this work is motivated by the following question: *"What level of individual consumer behavior inside a retail store can we reliably infer, by appropriately mining the sensor data from readily-available personal smartphone & smartwatch devices, without requiring ANY store-level infrastructural support"?*

While some of the high-end stores may already have WiFi infrastructure and in-store cameras, we emphasize that fact that our primary target is to provide a solution for the low-end stores in emerging economies. There can be cheap infrastructure such as BLE beacons that can be utilized; alternately, in the long run, maybe cloud-operated networked cameras become a cheaper and widely deployed option. However, using just video-based analytics, it is hard to accurately track an individual person (e.g., their clothing may change across different days) and personalized profiling of shoppers may prove difficult. As mentioned earlier, in this work we intend to profile individual shoppers and identify their fine-grained shopping behaviors. Moreover, our proposed solution has the advantage of being privacy-compliant, as the requisite sensor data is first captured on a user's mobile device and thus needs her explicit or tacit consent.

Driven by this objective, I present *IRIS* (**I**n-store **R**etail **Insights** on **S**hopper), an infrastructure-oblivious, mobile-cum-wearable based framework for in-store behavioral analytics of shoppers. *IRIS* is motivated by two key hypotheses: (i) A significant fraction of in-store browsing activities involve gestural interactions with objects of interest (such as picking up an item in a grocery store, retrieving and draping on a dress in a clothing store or having a coffee in the middle of a shopping episode), that a wrist-worn smartwatch should help capture; and (ii) A consumer's interest-level or familiarity level with objects of interest will also be manifested in macroscopic locomotion-related features (e.g., how long a person stood

Table 2.1: Taxonomy of attributes affecting individual consumer behaviors inside physical retail stores

Reference	Attributes	Examples of Shopping Behavioral Inferences
[113], [12]	Locomotive patterns	Store coverage, Time spent in store
[113], [12]	Gestural Interactions	Basket size or number of items purchased
[113]	Demographic characteristics	Gender-wise differences (e.g., item inspection times)
[74]	Cognitive state	Shopping intention (e.g., browsers vs buyers), Decision making styles
[74]	Interaction with personal devices	Online browsing behavior
[12]	Response to sales promotion devices/materials	Impulse buying, Preferred mode of intervention
[12]	Method of purchase	Mode of payment, Comes alone or accompanied by friends
[65], [113]	Store-level attributes (layout, location)	Familiarity with store, Crowdedness
[12]	Longitudinal characteristics	Heterogeneity of shopping trips, Store familiarity, Bulk shopper

stationary in front of a product), that a smartphone can help sense. Accordingly, we believe that a combination of smartphone & smartwatch sensor data can provide unique, hitherto unexplored, behavioral insights about a consumer's in-store behavior. While in this work our focus is on detecting *shopping gestures* and *macroscopic locomotion* (which contributes the main physical activities while shopping) of consumers inside a retail store, there are also additional aspects (see Table 2.1 for a taxonomy of individual consumer behaviors compiled based on several prior works in marketing and retail literature) that help to determine people's overall behavior in a retail store. For example, understanding a shopper's *cognitive state* during shopping activity would help in identifying their shopping intentions (e.g., whether shopper has a buying intent or not) [106]. Similarly of interest is to capture a shopper's interaction with their personal smart devices during shopping. Capturing such information would be helpful to obtain additional insights such as whether she is comparing the price of a specific item online or is she checking the shopping list to ensure that she has got all the items needed.

We explore the use of the *IRIS* framework to understand different aspects of individual-level behavior inside *retail grocery stores*. A key contribution of our research lies in appropriately *decomposing* an entire store visit (called a "shopping episode") into a series of modular and *hierarchical* individual interactions, such as a sequence of "in-aisle" durations, interspersed with "non-aisle" activities. Each "in-aisle" segment can consist of one or more product-interaction activities, such

Figure 2.1: Typical sequence of shopper activities in a grocery store

as “picking up item” (**P**), “putting item in trolley (cart)” (**T**), or “putting item back in the aisle” (**B**). Figure 2.1 visually illustrates such a decomposition. This decomposition is crucial because it not only helps define the specific atomic “activities” for which we seek to extract discriminatory features and build classifiers, but also helps to conceptualize two different levels of individual-level behavior (these will be further detailed in Section 2.2).

IRIS’ broader vision (not explored in depth in this dissertation) is to use longitudinal observations (across multiple shoppers) of such individual-specific, episode-level insights, to infer both: (a) Store-Level Properties, such as whether a store’s layout is confusing or if the store lacks the right selection of products (a large number of shoppers have unproductive shopping episodes) and (b) Individual-Level *Persona*, such as whether the shopper is always in an hurry while shopping, or how often she purchases unfamiliar items or visits unfamiliar stores (an indication of her level of adventurousness).

Even in absence of item-specific knowledge, such insights can enable new applications such as: (a) *targeted advertising:* e.g., promotions of newly launched products preferentially pushed to shoppers whose prior browsing behavior indicates a propensity to look for unfamiliar products (so-called diversity-seeking behavior); (b) *proactive retail help*: e.g., a shop assistant directed to assist the customers who exhibit an “undecided” purchase pattern (an unusually high number of items picked up from, but then returned to, the shelves); or (c) *crowdsourced store profiling: IRIS* can be built as a 3^{rd}-party mobile App, as it does not have any interaction with the store's IT infrastructure. Accordingly, crowdsourced data from a pool of shoppers using *IRIS* can be used to build typical "experience profiles" associated with the store, for use in recommendation applications. There is a slowly increasing trend of consumers downloading and using apps that utilize mobile sensing to enhance their shopping experience [6, 4]. For example, *Appadia Mall App*, *Bleesk App*, *NearBee App* [6] are some of the retail apps that have become recently popular and being used by shoppers.

2.1.1 Key Challenges and Research Questions

IRIS' broad goals require us to address several research questions:

1. *Shopping Interaction Recognition:* Given sensor data corresponding to a specific shopping gesture (e.g., putting an item in the cart), what discriminative features help us identify such gestures? What level of accuracy for individual-level gesture recognition can we achieve, by intelligently combining sensor data from both smartphones and smartwatches?

2. *Accurate Episode Segmentation:* Given that a shopping episode can consist of a shopper's interaction with multiple items, and movement across multiple aisles, how do we take the sensor data for the entire episode duration and then reliably segment it into individual interaction instances (such as in Figure 2.1)? What are the errors in demarcating the (start, end) times of such individual interactions?

3. *Connecting Interaction-Level Observations to Overall Behavior:* Assuming that we can infer the individual-level interactions of a shopper (i.e., how many items the shopper placed in her cart, etc.), how reliably can we use such inferences to classify the overall episode-level behavioral attributes (such as

whether a shopper was in a hurry or not)? Can such classification be person-independent, or do shoppers behave differently enough to warrant person-specific classifiers?

We address these questions, by utilizing a fairly extensive set of user studies (detailed in Section 2.3), involving 50 distinct shopping episodes, collected from 25 individuals, across 2 different mid-sized retail grocery stores in Bengaluru, India. Overall, by answering these questions, we show how *IRIS* can accurately recognize various aspects of in-store shopping activities and obtain fine-grained insights on shopper's item-level interactions, episode-level attributes and store-level attributes.

2.1.2 Key Assumptions and Limitations

As several capabilities can be desired in a smart retail application, we establish upfront the assumptions taken in *IRIS* and the functionalities that *IRIS* does not currently support. Very specifically, we make the following assumptions:

- The shopper wears a smartwatch in their dominant hand and carries a smartphone in their pocket.
- The shopper performs the item-interactions with his/her dominant hand.
- The shopper picks only a single item at a time and puts into the trolley before interacting with the next item.
- The shopper puts the item into the trolley within the 'aisle' itself. When in 'non-aisle' zones, the shopper is pushing the trolley and walking.
- There is no active engagement with consumers in non-aisle areas–e..g, there are no 'live demonstrations' by in-store employees which may affect the shopper's browsing and in-store navigation pattern.

While *IRIS* supports novel capabilities such as inferring shopper's item-level interactions and building a personalized shopper profile, it currently has the following limitations:

- *Support for only three shopping gestures:* Currently, *IRIS* detects only "picking", "putting an item back" and "putting an item into a trolley" gestures. However, the capabilities of *IRIS* can be extended in future to recognize other shopping gestures such as "inspecting an item", "trying out an item".

- *Tracking of only gestures performed with dominant hand:* As *IRIS* assumes that the shopper wears a smartwatch only on their dominant hand, it can only track gestural interactions performed by that arm. Therefore, in its current state, *IRIS* fails to capture actions performed by the alternate arm and also has limitations in fully supporting multi-arm gestures.

- *Does not identify exact item being picked or interacted:* As *IRIS* works without the support of any location-based technologies, it cannot identify the exact item that is being picked or interacted by the shopper. However ultimately, when it is combined with such positional or product arrangement information, we can more accurately capture specific items interacted by the shopper and enable applications that can provide item-specific deals and reminders.

2.2 IRIS: Architecture and Key Objectives

IRIS' goal is to uncover shopper-specific and store-level behavioral attributes, both during a specific shopping episode, and via aggregated observations across a longitudinal trace of such episodes. As *IRIS* does not presuppose any support from the store (e.g., location tracking, maps, PoS data, etc.), it does not attempt to capture insights such as specific product viewed or bought by a shopper. Instead, our goal is to infer *item-independent* aspects of a shopper's behavior, such as number of products picked and then returned, movement speed within the store etc.

2.2.1 Types of individual and store-level insights

One of our long-term goals is to use microscopic *gestural-level* insights obtained during a consumer's interaction with a single product as a "building block", to help build progressively deeper insights about both a shopper's short-term and longer-term behavioral attributes. In this view, the item-specific insights gained by looking at a set of sensor data *frames* (a relatively small duration lasting a few seconds) can be viewed as elements of a periodic table of in-store shopping behavior; these elements are then combined in hierarchical fashion to discover the higher-level individual and store-level attributes. More specifically, we categorize the insights into three broad bins:

• *Item-Level Insights (Individual):* These insights describe aspects of an individual shopper's behavior with a specific product (or product type). For example, based on the time that the user inspects the product, i.e., the interval between a 'P' (pick) and the corresponding 'T' (in trolley) activity, we hope to learn if this is a "familiar" product (that the shopper regularly buys without much additional thought) or an "unfamiliar" one. Similarly, an observation of multiple 'P' (picks) and 'B' (put backs), before an eventual 'T' (trolley), might indicate that the shopper had no a-priori brand affinity, but instead compared multiple brands before picking a specific item. These item-level insights are derived primarily based on the *"gestural interactions"* attribute, outlined in Table 2.1.

• *Episode-Level Insights (Individual & Store):* These insights are obtained at the shopping episode-level (an episode comprises multiple item-level interactions) by aggregating individual item-level labels/features. They are inferred based on a combination of consumer behavior attributes such as *locomotive patterns* and *gestural interactions*. These insights can capture the episode-level behavior of the shopper (e.g., a relatively small number of in-trolley ('T') actions, coupled with shorter "non-aisle" durations, might indicate that the "shopper was in a hurry"). Moreover, the insights can also describe properties of the store itself (i.e., *store-level* *attributes*)–e.g., unusually slow movement during "non-aisle" segments might indicate that the store was overly crowded.

• *Longitudinal Insights (Individual & Store):* These insights are obtained by aggregating observations across a large collection of episodes (independent store visits), observed over a period of weeks and months. These *longitudinal characteristics* help to better understand the behaviors of individual shoppers as well as the properties of a store. At an individual-level, they can help reveal the shopper's *persona*— for example, that the "shopper is always hurried during a weekday visit" or that "the shopper always shops in bulk". At a store-level, they can help reveal the store's macroscopic properties — for example, that "store X has more (or less footfall) during specific times or days".

In this work, given the absence of longitudinal data, we focus only on item and episode-level behavior of shoppers.

2.2.2 The IRIS Architecture

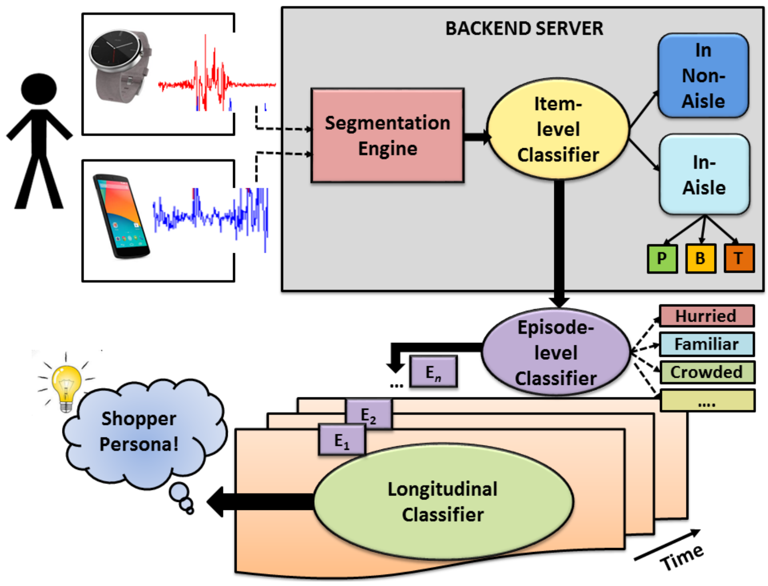


Figure 2.2: Functional Components & Analytics Flow

Figure 4.1 illustrates the device and backend components of the *IRIS* framework, as well as the typical flow of the analytics pipeline. Each individual shopper carries an on-body smartphone and smartwatch, whose sensor streams capture the individual's physical movement and gestural activities, over an entire shopping episode. At the backend, this entire stream is first run through a *Segmentation Engine*, which splits up the entire shopping episode duration into different segments (time chunks), each corresponding to a single movement or gestural activity. Each individual chunk is then fed into a hierarchical *"Item-level"* classifier, which attempts to first classify each chunk as either "in-aisle" vs. "non-aisle", and subsequently separately classifies different gestures within an "in-aisle" segment into one of multiple interaction-related labels (e.g., {P, B, T} gestures). This collection of gestures and movement patterns (from the Item-level classifier) is then collectively analyzed by the *Episode-level* classifier, to help discern episode-level labels (e.g., "was the shopper in a hurry?"). Finally, the *Longitudinal Classifier* operates at longer time scales, analyzing (a) multiple episodes of the same shopper to determine *"persona-level"* attributes, and (b) episodes from multiple in-store shoppers to determine *"store-level"* attributes.

2.3 Dataset

We first describe our process of collecting real-world shopping behavioral data. We conducted a user study with 25 middle-aged volunteers (15 females, 10 males) recruited from Xerox Research Centre, located in Bengaluru, India [1]. The study was conducted during the period of July-August, 2015. Each participant was asked to visit two different retail grocery stores in Bengaluru, India (one large and spacious, the other much more cramped for space) and purchased items from a given shopping list. We collected 50 shopping episodes from the grocery stores at different times of the day. Each episode lasted, on average, for about 20 minutes and belonged to one of 3 distinct types: (i) *Engineered List* (20 episodes), (ii) *Clocked* (20 episodes) and (iii) *Discretionary* (10 episodes).

[1] Xerox Research Centre, now Conduent Labs, India – (`https://www.conduent.com/innovation/`)

Engineered List: The participants were given a list of 14 grocery items which consisted of 4 Frequent-Choice (*FC*), 4 Infrequent-Choice (*IC*), 3 Frequent-Specific (*FS*) and 3 Infrequent-Specific (*IS*) items. The items were categorized based on general consensus after a small survey. For example, egg and bread were frequent items, while dish-washing soap and Schezwan sauce were infrequent items; "select a juice of your choice" is an example of a FC item; while "Tropicana Orange Juice– 1 gallon" exemplifies a Specific item. The participants were asked to shop for the items in the same order as in the list.

Clocked: The objective here was to emulate "hurried" behavior. Hence, we paired up the participants, gave each a list of 10 items and engaged them in a shopping competition. The participants were informed that the person clocking the least overall time, while buying all the items listed, would be declared the winner. To control for differences in familiarity with the shop, all the participants in this experiment were familiarized with the shop and its aisles before the episode started. All items in the list were open-ended (Choice), and selected "randomly" (by picking ingredients from arbitrary common recipes).

Discretionary: The objective here was to capture behavior in situations where a shopper could choose not to buy an item, due to a variety of factors (such as budget constraints, product unavailability, or deficient quality). The items in the list were chosen to elicit some of these factors. Sample items included fruits that were out of season, items with budget constraints which were not feasible, "greens that needed to be fresh enough", "red coffee mug with a design they liked", etc. These are cases of 'discretionary purchases', that depend on the shopper's qualitative assessment. The shoppers were unaware of our study objectives; the traces thus capture the natural behavior of shoppers who earnestly look for a preferred item but may be unable to find it.

Table 2.2: List of sensors monitored

Sensors	Purpose	Device
Accelerometer	Speed and patterns in walking and hand activities	Watch, Phone
Gyroscope	Rotational and angular information during walking and hand activities	Watch, Phone
Magnetometer	Directional information during walking	Phone
Step Counter	Number of steps directly obtained from Google Fit API	Watch, Phone
Battery Temperature	Distinguish between zones using ambient temperature (e.g., freezer section)	Watch
Light Sensor	Ambient lighting in the store	Watch
Audio Sensor	Ambient noise in the store	Watch
Heart Rate	To study if specific browsing behavior causes excitement	Watch

2.3.1 Sensor Data Collection

Each participant was given a smartphone (running Android v4.3 or above) and a smartwatch (Android Moto 360). The phone was placed in the right-side pant pocket facing front, and the watch was worn on the dominant hand (all our participants were right-handed). The devices were pre-installed with our custom data collection apps for the smartphone and smartwatch. The apps recorded data from the sensors listed in Table 2.2, at the maximum permitted sampling frequencies of 200 Hz (phone) and 25Hz (watch). Some sensors were only exclusive to a single device–e.g., the magnetometer was unavailable on the watch, whereas the heart rate sensor was unavailable on the phone. Ambient sensing (temperature, light and audio) was more reliable on the watch since the phone was placed inside the pocket.

2.3.2 Ground Truth Collection

The ground truth of a shopping episode was collected by having a person *shadow* the shopper (without the shopper's knowledge). The shadower used an app on his own device, which enabled him to both record micro-activity labels of the shoppers ("Picking", "In Trolley", "Enter Aisle", etc.), and to record audio notes, along with the timestamps (all three devices, i.e., shopper's phone & watch, and shadower's phone, were time-synchronized). Other non-activity related information, such as the shopper's *familiarity level* with the store or the *crowdedness* of the store were captured via a survey filled in at the end of each episode. To ensure uniformity in ground truth annotation, an item-level interaction was assumed to start after the preceding "Trolley" label (where the user was pushing a trolley), and continued till the subsequent "Trolley" label; the interval itself could contain multiple labels, such as "pick", "put back", etc. Note that all our studies (and analyses) make the assumption that the shopper always uses a trolley, although we believe that the technique can be extended to other modes (e.g., a shopping basket). We also omit the case whereby the person does not use a trolley or a basket, since then the person would not be buying many items and consequently the store operator may have lower interest in analyzing such behavior.

2.4 Classifying under Perfect Segmentation

As the first step in investigating *IRIS*, we first seek to extract the discriminatory features of smartwatch & smartphone sensors, and understand their classificatory power, to help infer various shopper-experience related item-level and episode-level properties. More specifically, in this section, we assume that, via some as-yet unknown mechanism, we have perfect knowledge of the (start, end) times of each item-level interaction (e.g., the "P", "B", "T", "in-aisle" or "out-of-aisle" activities), and investigate two questions via a supervised classification approach: (1) How accurately can we classify each of the distinct item-level interaction activities, and what features aid this classification? (2) Given knowledge of such item-level behavior, how accurately can we infer *episode-level* properties, and what features (defined over the aggregated item-level interactions) aid this classification?

Table 2.3: Features for Item-level classification

(1)	Mean number of picks
(2)	Variance in number of picks
(3)	Mean hold time, i.e., duration between picking and putting back
(4)	Variance in hold time
(5)	Mean duration of time between picking an item for the first time and putting in trolley (W1)
(6)	Variance in W1
(7)	Mean Duration between entering an aisle to putting item in trolley (W2)
(8)	Variance in W2
(9)	Mean Duration between walking in non-aisle to entering an aisle (W3)
(10)	Variance in W3
(11)	For each time window W1, W2 and W3, following features from phone accelerometer: mean & variance in magnitude, spectral entropy & energy.

2.4.1 Item-level Shopper Experience Attributes

We start by trying to identify the following four item-level attributes (based on the shopper's interaction with that specific item), as insights on these four attributes help reveal a shopper's buying preferences and habits: • *Frequent Item:* An item that the shopper buys frequently or routinely and is familiar with. • *Infrequent Item:* An item that the shopper is less familiar with because he does not buy it as often. • *Specific Item:* An item for which the shopper has *a-priori* knowledge of the specific brand & product detail. • *Choice Item:* An item for which the shopper does not have an *a-priori* product in mind, but instead needs to view alternative products and make a choice.

Table 2.3 lists the various features that we used to classify these 4 labels. The features have a hierarchical structure as follows. Initially, different statistical features (similar to that used in [135]) are used to identify each interaction/movement activity as "P", "B", "T", "in-aisle" and "non-aisle". While the phone-based features help identify the walking/gait-related patterns (e.g., "in-aisle" or non-aisle), the watch-based features help identify the gestural interactions ("P", "B", "T"). Subsequently, features (1-10 in Table 2.3), defined over the interaction and movement activities, help classify the *item-level* aspects of shopper experience.

Features 1-10 were defined to help exploit several intuitive properties of human behavior that we visually observed across shopping episodes. For example, for either a Specific (*S*) or a Frequent (*F*) item, we can expect the shopper to perform a smaller number of picks (P), exhibit smaller hold time (H), as well as have smaller durations of the time windows W1, W2 & W3. In contrast, for Choice (*C*) or Infrequent (*I*) items, shoppers will likely exhibit a larger number of pick (P) and put back (B) gestures and a longer duration of window W1 (as they evaluate multiple items before converging on a selection). Moreover, for Infrequent items, shoppers will likely spend more time and effort to locate the item, resulting in larger durations of windows W2 and W3. Note that the analysis of *F* vs. *I* is performed by considering only those users who were familiar with the store, to avoid the confusion on whether a shopper's item-level behavior was due to unfamiliarity with the item or the store's layout.

Figure 2.3 shows the values of these features for each of these classes averaged across all episodes we collected, in order to gain insight into the dataset w.r.t these features. We see that the data reflects certain intuitive or expected trends. For example, compared to *S* items, *C* items have a higher mean duration for windows W1 and W2 (features 5 & 7); similarly, *I* items tend to exhibit longer durations of non-aisle movement (feature 9). To understand the ability of these features in classifying these product-level attributes, we trained J48 decision tree classifiers, along with Correlation Feature Selection (CFS) to identify the most dominant (discriminatory) features. Note that we trained 3 different classifiers, two binary classifiers (one each to distinguish between *S* vs. *C* and *F* vs. *I*) and one quaternary classifier (to distinguish between the 4 composite labels (*FS*), (*IS*), (*FC*) and (*IC*)).

Table 2.4 tabulates the results obtained via 10-fold cross validation. We note that we get almost 100% accuracy (both precision and recall values are over 99% for all labels)! This is a very encouraging result, especially given that our dataset contains labels aggregated from 25 users, who we expect have diverse shopping styles and preferences. **These results suggest that the behavioral markers of shoppers are**

Table 2.4: Item-level classification with ground truth. Column 3 uses indices from Table 2.3

	Precision	Recall	Dominant Features
Frequent	0.997	1.0	(1), (3), (5),(6)
Infrequent	1.0	0.998	(1), (3), (6)
Specific	1.0	0.999	(2), (1), (4), (5)
Choice	0.999	1.0	(2), (1), (4), (5)
Freq-Spec	0.993	0.999	(2), (1), (4), (5)
Freq-Choice	1.0	1.0	(2), (1), (4), (5)
Infreq-Spec	1	0.997	(2), (1), (4), (5)
Infreq-Choice	1	0.998	

distinct enough (between $\{S, C, F, I\}$ products) for us to robustly identify them from a combination of smartwatch and smartphone sensor data.

To study if there was any gender-specific differences in the shopping behavior and the respective results, we conducted the analysis separately for males and females in our dataset. However, we did not observe any notable difference in the results across the two gender categories. Given that we had only data from 25 shoppers, further extended studies with more number of participants (also of varied age groups) and longitudinal studies with the same shopper are required to indeed confirm if any differences exists in the shopping patterns across males and females and its impact on *IRIS* technologies.

2.4.2 Episode-level Shopper Experience Attributes

We next focus on inferring the individual-specific episode-level characteristics, such as whether the shopper was in a hurry (or not) or whether the shopping experience was productive (i.e., did the shopper find most of the items he was looking for?). Following the approach used previously, we used J48 binary classifiers and features 1-10 (listed in Table 2.5) to study whether a shopper was "hurried" or not. The data from 20 hurried ("Clocked") episodes were combined with 20 non-hurried ("Engineered List") episodes to perform the *HU* vs. *NH* analysis. Figure 2.4 shows the values of these features averaged across all episodes from our data set, correlating

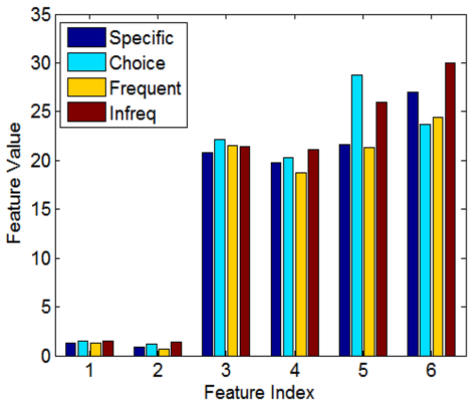


Figure 2.3: Values of dominant item-level features listed and indexed in Table 2.3

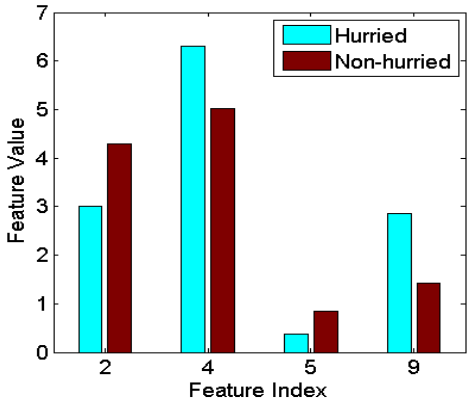


Figure 2.4: Values of dominant episode-level features listed and indexed in Table 2.5

as expected with the feature set. With these features, a J48 decision tree binary classifier yielded an overall precision and recall of 99% each, which are tabulated later in Table 2.8 for comparison. Features (2), (4) & (9) were the most dominant.

2.4.3 Store-level Shopper Experience Attributes

We finally focus on shop-level attributes–i.e., conclusions that we can derive about the shopper's perception or interaction with the overall store. Given our dataset, we focused on one specific store-level attribute (as this could be corroborated from the post-episode survey data): was the shopper familiar (*FL*) or unfamiliar (*UFL*) with the store? For this study, we only utilized data from the "Engineered List" and "Discretionary" episodes. We ignored the "Clocked" episodes as it simulated a hurried behavior and the shoppers were also familiarized with the store layout. Out of these 30 episodes, 11 reported they were "unfamiliar" with the store. For evaluation, we used a balanced set of 11 randomly chosen *FL* and 11 *UFL* episodes of data.

We observe that a shopper's non-aisle behavior is most indicative of his store-level unfamiliarity: unfamiliarity results in increased effort (time) in trying to locate the correct aisle, and the item within the aisle. The feature set in Table 2.5 captures this behavior: as seen in Figure 2.5, the average values of these features are quite

Table 2.5: Feature set for determining hurriedness & store familiarity of shopper (* marks the dominant features)

Attribute	Features
Hurried/ Non-hurried	(1) Mean duration in an aisle, (2)* Variance of duration in an aisle, (3) Mean duration in a non-aisle, (4)* Variance of duration in a non-aisle, (5)* Mean step rate in an aisle, (6) Variance in step rate in an aisle, (7) Mean step rate in a non-aisle, (8) Variance in step rate in a non-aisle, (9)* Mean hold time, (10) Speed of picking an item (mean magnitude of watch accelerometer during a pick)
Familiar/ Unfamiliar	(1) Total time spent in the shop normalized by the number of items, (2)* Step rate in non-aisle, i.e., $\frac{number of steps}{duration}$, (3) Fraction of time spent in non-aisle, (4)* Mean of duration from entering an aisle to first pick, (5)* Std deviation of duration from entering an aisle to first pick, (6) Mean step rate from entering an aisle to first pick, (7)* Std deviation of step rate from entering an aisle to first pick

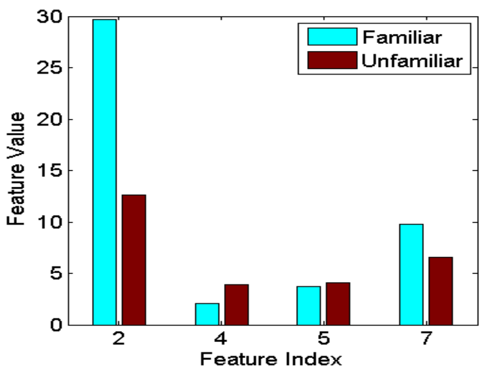


Figure 2.5: Mean values of features (across 22 episodes) listed and indexed in Table 2.5. All durations are in seconds

distinct between *FL* & *UFL* users. As before, we use a J48 decision tree binary classifier with this feature set, correlation feature selection and 10-fold cross validation. We achieve 99.99% precision and recall (with (2), (4), (5) and (6) being flagged as the dominant features), indicating that we are able to very reliably distinguish between *FL* and *UFL* shoppers.

Summary: Our results in this section indicate that *IRIS* can indeed very reliably (with accuracies usually above 99%) infer item-level and episode-level aspects of a shopper's in-store behavior. However, there is a big caveat: our high accuracy has been demonstrated (thus far) only under the assumption that the overall sensor data has been reliably *segmented*–i.e., the (start, end) times of each activity label are correctly known. We next develop novel techniques to perform such automated and accurate segmentation.

2.5 Automatic Segmentation

The supervised learning discussed in Section 2.4 assumed the use of ground truth labels to demarcate the time segments corresponding to different activities. Another key contribution of this work, we now describe how to automatically deduce the (start, end) times of various labels through a combination of (i) landmarking based on significant sensor features, to distinguish between non-aisle and aisle zones (ii) Viterbi decoding to predict the sequence of hand activities and (iii) improving the precision of this hand sequence prediction by estimating the likelihood of an item being found using survival analysis models, and utilizing this information to bias the transition probabilities in a time-dependent markov model.

2.5.1 Differentiating Aisle and Non-aisle zones

The key observation used in landmarking aisle and non-aisle zones is that when a shopper moves into an aisle to look for an item, there is a marked difference in the walking speed and hand movement, as he slows down after entering an aisle of interest. The inter-step interval (i.e., the duration between consecutive steps) is higher inside an *aisle* than in *non-aisle* (*note:* this assumption is valid if the aisle is non-crowded); moreover, while a shopper mostly pushes the cart (or carries a basket) in non-aisle, he has a lot more variations in the hand movements due to various browsing-related actions. Further, the inter step interval for a shopping episode (Figure 2.6) reveals that an *aisle* zone always begins from the foot of a peak until the peak; similarly, a *non-aisle* zone spans from the peak to the foot. However, the number of peaks spanned, i.e., duration for each zone is variable. Accordingly, using peak and valley detection, we identify all peak-points

Table 2.6: Feature set for classifying aisles/non-aisle zones and hand/non-hand activities

Feature	**Aisle vs Non-Aisle**	**Hand vs Non-Hand**
Mean phone accelerometer magnitude	✓	✓
Spectral entropy of phone accelerometer magnitude	✓	✓
Mean watch accelerometer across x,y,z axes	✓(only y,z axes)	✓
Spectral entropy of Watch accelerometer across x,y,z axes	✓(only y,z axes)	✓
Mean watch gyroscope along x,y,z axes	✓(only x-axis)	✓
Spectral entropy of watch gyroscope along x,y,z axes	✘	✓
Variance in step rate	✓	✘

(t_{peak_i}) and foot-points (t_{foot_i}) of all ramps. In order to determine the duration of the zones, we use change point detection analysis using a binary random forest classifier trained to identify *aisle* and *non aisle* regions, using statistical features from phone accelerometer, watch accelerometer and watch gyroscope listed in Table 2.6. A sliding window size of 10 seconds was used. The precision and recall of this classifier model is 0.888 and 0.875, respectively. The reasoning behind the change point detection algorithm is that the classification probability will drop when the test set contains mixed data, i.e., data from across different categories. Accordingly, we first gather the features within the window corresponding to the first ramp, $w = [t_{foot_1}, t_{peak_1}]$ and compute the probability $Pr(aisle|featureset(w))$ using the binary classifier. Next we increase the window size to include subsequent peaks, one peak at a time, until the classification probability drops. Suppose the accuracy dropped for the window $[t_{foot_j}, t_{peak_i}]$, the region $[t_{foot_j}, t_{peak_(i-1)}]$ are marked as "aisle". Similarly, next the features in window $w = [t_{peak_(i-1)}, t_{foot_i}]$ is used to compute $Pr(nonaisle|featureset(w))$, and the window size is incremented to include subsequent troughs in the acclerometer data until the probability drops, say at t_{foot_k}; the region $[t_{peak_(i-1)}, t_{foot_(k-1)}]$ is then marked as "non-aisle".

Accuracies for segmenting aisle and non-aisle regions are as shown in Table 2.7.

The possible reason for higher false positives in classification of non-aisle is because of the "walk-and-browse" characteristic, i.e., the time instances when a shopper continues to walk after entering the aisle, without necessarily slowing down or picking items to check items. The average offset in time between an actual segment and predicted segment is around 5 seconds.

2.5.2 Identifying Hand Activities

There are two parts to solving the problem of identifying hand activities, which are defined as either a *Pick* (P), *Put Back* (B) or *In Trolley* (T). The first is to identify if any hand activity occurred, and if so, the next is to identify which of these three actions it was. The first part is straightforward by analyzing the gyroscope data from the smart-watch. Figure 2.7 shows the gyroscope data, after performing quaternion rotation with respect to a common origin [111], and fitting it to a spline curve [84]. The value plotted is the normalized product of pitch, roll and yaw. The figure also shows the ground truth in terms of the times when a hand activity did occur. We observe that the peaks are a good indicator of a hand activity, with negligible false negatives, but there are a significant number of false positives, resulting from arbitrary hand movements. To address this, we first run a peak detection algorithm to identify the peaks and then eliminate bulk of the false positives by filtering out those peaks that occur during *non-aisle* segment (as described in Section 2.5.1). For each remaining peak, we compute the features in the window corresponding to the width of the peak (full-width at half-maximum), and feed it to a random forest binary classifier to compute the probability that it is a hand activity based on a combination of watch gyroscope, watch accelerometer and phone accelerometer features (Table 2.6). This process yields a precision of 95% and recall of 98% in identifying a hand activity.

The next step after identifying the existence of a hand activity, is to predict if it is a P, B or T. We propose using a Viterbi decoding approach on a Hidden Markov

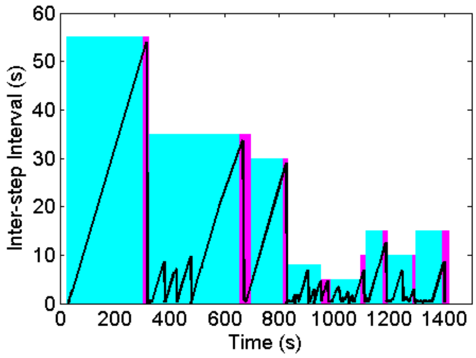


Figure 2.6: Inter-step interval and corresponding aisle (dark blue) and non-aisle (light pink) zones

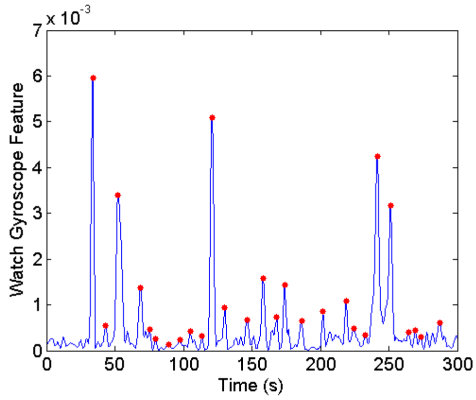


Figure 2.7: Watch gyroscope peaks indicating potential hand activities. The red dots show the actual hand activities from ground truth.

Model in order to leverage the inherent sequential nature of gestures in a shopping episode. The state transition probabilities between P, B and T are computed from the experimental data. The trellis diagram corresponding to the Viterbi decoding is shown in Figure 2.8. The emission probability is defined as $Pr(FS|l)$, where $l = P, B, T$, and $FS = [f_1, f_2 ... f_n]$ is the set of features from watch gyroscope and watch accelerometer (features 3, 4, 5, 6 in Table 2.3), which are the observations in our HMM. The emission probability is obtained as:

$$Pr(FS|l) = \frac{Pr(l|FS) * Pr(FS)}{Pr(l)}, \tag{2.1}$$

where $Pr(FS) = \prod_{i=1}^{n} Pr(f_i)$, since sensor features are independent. The probabilities $Pr(f_i)$ and $Pr(l)$ can be obtained from the distribution of the empirical data. The probability $Pr(l|FS)$ is obtained from the random forest ternary classifier, which is trained to distinguish between P, B and T using the features in FS (with an average precision and recall of 0.926 and 0.927 respectively).

One salient aspect about this decoding approach is that it avoids onset of cascaded prediction failures. This is because, the length of the predicted sequence is limited to each aisle segment, i.e., the sequence is predicted independently for each aisle segment, since the activities within each aisle-segments are independent

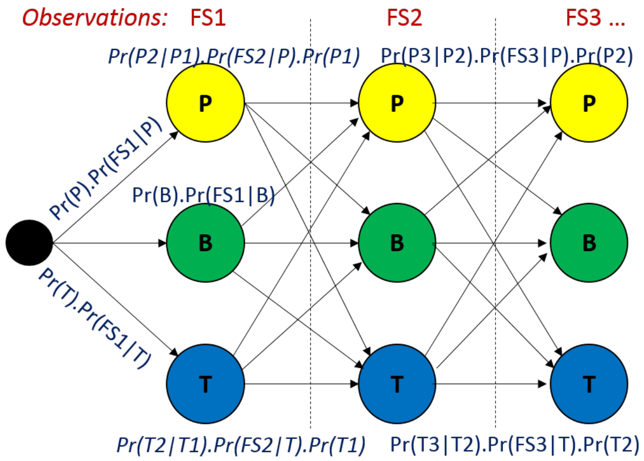


Figure 2.8: Trellis diagram corresponding to the Viterbi decoding of hand action sequences.

of other segments, and this helps contain prediction errors. The performance of classification is shown in Table 2.7.

Table 2.7: Accuracy of automatic segmentation in identifying Aisle, Non-aisle, Pick, Put-Back and In-Trolley actions

	Aisle	Non-aisle	P	B	T
Precision	0.9775	0.9051	0.9863	0.9149	0.8200
Recall	0.9669	0.9376	0.9863	0.9053	0.8367

2.5.3 Survival Analysis

We see that the prediction accuracy for T is lower than the other two activities, and is often mispredicted as B. In order to improve the accuracy, we use the likelihood of finding the item as an indicator of whether the action would converge in a Put Back or a Trolley. This probability is obtained by using the Cox Proportional Hazards model [22], which is a semi-parametric method for adjusting survival rate estimates to quantify the effect of predictor variables. We chose this model based on the intuition that *Trolley* actions takes longer, therefore, longer the duration, higher the chances of T. Accordingly, we estimate the likelihood that an item is found, given the time taken since the search began (i.e., the last aisle entry or T event, whichever comes later), with the number of picks as a covariate, i.e., an explanatory or predictor variable, that may affect the survival time. Let T be a continuous random variable representing the waiting time until an item is found. Let $f(t)$ be the probability density function of T and $F(t) = Pr\{T < t\}$ its cumulative distribution function. The survival function $S(t)$ is then defined as the probability that it takes more than time t to find an item, or in other words, the probability that the failure event of not finding an item has not occurred by duration t. This is given by:

$$S(t) = Pr\{T \geq t\} = 1 - F(t) = \int_t^\infty f(x)dx \tag{2.2}$$

The hazard rate function $\lambda(t)$, which is the instantaneous rate of occurrence of the failure event, is defined as:

$$\lambda(t) = \lim_{dt \to 0} \frac{Pr\{t \leq T < t + dt | T \geq t\}}{dt} = \frac{f(t)}{S(t)} = -\frac{d}{dt} logS(t) \tag{2.3}$$

According to the proportional hazards model, we have $\lambda_i(t|x_i) = \lambda(t)e^{x_i \beta}$, where x_i is the vector of covariates, β is the coefficient of x_i. It is assumed in this model that the covariates remain constant over time, however the number of picks is not constant. Hence, we treat each discrete value of number of picks as a separate covariate, derive a different hazard function for each case, i.e., we derive λ_i for $x_i = 0, 1, 2...8$, where the maximum number of picks for a single item that we have in our data set is 8, and subsequently obtain the corresponding survival functions. Our analysis shows that the family of survival functions obtained this way has 81.3% accuracy in predicting the likelihood of an item being found. When it is time to predict a label for a hand action, we determine the duration τ of the current item-episode, i.e., the time between the latest Aisle (or last T) and the start of the hand action segment, and the number of picks x during this time (as predicted). From the survival function corresponding to x, with τ as its input, we obtain the likelihood of the item being found.

If the item is likely to be found (> 0.5), then we bias the sequence prediction towards a T, or else towards a B. The biasing method is different in each case. In order to bias towards a Put Back, we multiply the state transition probabilities for the transitions into T by $S_x(\tau)$. In order to bias towards a T, we make use of the fact that as the number of B for an item increases, and the item is likely to be found eventually, the likelihood of a T increases, i.e., there arises a time-dependent Markov chain. We retain the Markov property by conditioning the states based on the number of prior Pick-Put Back actions during that item-episode. In other words, we compute different multiple transition probability matrices where the probabilities for transitions are conditioned based on the number of prior Pick-Put Back transitions seen in the current item-episode. In other words, we compute a family of transition probability matrices $\{TPM_i\}$, where each matrix TPM_i gives the transition probabilities between Pick, Back, Trolley given there have occurred i Pick-Put Back prior transitions. Every time the state B is entered from P, a counter i is incremented by 1, to index into (and hence use) the i^th transition matrix TPM_i. Every time state Trolley is entered, the counter is reset back to 1 (for next item). **Using this approach the precision and recall of prediction of T improved by** 7.6% **and** 4.2%**, respectively. to 0.8830 and 0.8646, respectively; the precision and recall of prediction of B improved to 0.9226 and 0.9337**.

2.5.4 Attribute Classification with Automatic Segmentation

Finally, we re-ran the supervised learning classification experiments described in Section 2.4, with the same set of features, but with labels obtained from our automatic segmentation approach instead of ground truth. We compared accuracies of (a) classification with ground truth labels and (b) classification with a brute-force approach for automatic segmentation. The basic idea behind this brute force approach is to use a regular classifier to determine which label a time window belongs to. Accordingly, we split our data into windows of 10 seconds. We then use the binary classifier trained with the features in Table 2.6, as discussed in Section 2.5.1 to determine if each window belongs to *Aisle* or *Non-Aisle*. Next, for each predicted aisle segment, we split it into 3 second windows, compute the features in Table 2.3

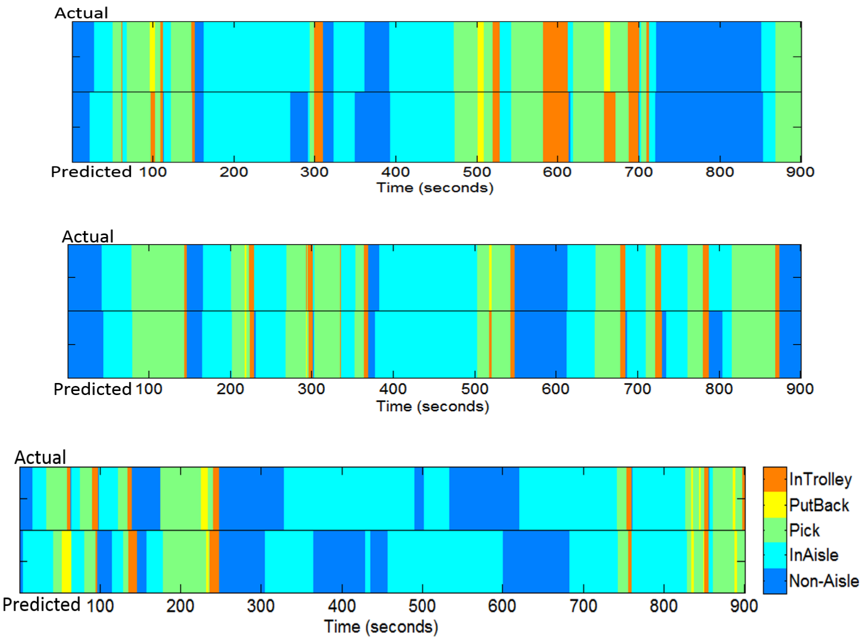


Figure 2.9: Trace of actual and predicted labels in the first 15 minutes of a shopping episode for (a) unhurried, familiar shopper (b) hurried, familiar shopper and (c) unfamiliar, unhurried shopper.

Table 2.8: Comparison of Classification Accuracy with different approaches

In this table, the columns, **P** and **R** represent the precision and recall obtained using our automatic segmentation approach, **P_gt** and **R_gt** represent the precision and recall obtained from ground truth labels, **P_bf** and **R_bf** represent the precision and recall obtained with brute force approach.

Attribute	**P**	**R**	**P_gt**	**Rgt**	P_**bf**	**R_bf**
Freq - Infreq	0.921	0.926	0.99	0.99	0.653	0.666
Specific-Choice	0.88	0.89	0.99	0.99	0.553	0.644
Familiar-Unfamiliar	0.92	0.92	0.99	0.99	0.644	0.721
Hurried-Non hurried	0.916	0.922	0.99	0.99	0.693	0.714

for these windows, and use the ternary classifier discussed in Section 2.5.2 to determine if that window belongs to a P, B or T. We decided these window sizes of 10 seconds and 3 seconds after some trial and error, selecting that window which gave the highest accuracy. This brute-force approach only yielded an average precision and recall for Aisle/Non-aisle of 71.3%, 73.5%, respectively; and for P, B, T classification, 50.5% and 38.8%,respectively.

Next we re-ran the attribute classification after automatic segmentation. Table 2.8 shows the average classification accuracy for item level and episode-level attributes, using our automatic segmentation method and brute force approach.

We see that our segmentation yields very good accuracy. Interestingly, we see that the accuracy is higher for Frequent vs. Infrequent and Hurried vs. non-hurried, than the other classification. This is most likely because the dominant features of these attributes involves non-aisle and picks which are more accurately predicted, than those that involve trolley and put back labels.

Figure 2.9a shows a sample trace of predicted and actual labels for the first 15 minutes of a shopping episode for a shopper who was not in a hurry and was familiar with the store; Figures 2.9b and 2.9c show similar traces for unhurried-familiar, and unhurried-unfamiliar shopper, respectively. Interestingly we can see that these traits of a shopper are revealed to some extent in the traces. For instance, a hurried shopper has fewer put backs and an unfamiliar shopper spends longer durations without interacting with items. We also observe that the classification accuracy of our framework varies with such profiling. For instance, the accuracy of Put Back and Non-aisle for hurried shopper is lower than average (87% and 84%), which can be reasoned that the gestures are performed in a hurry, and the shopper walks fast in both aisle and non-aisle when in a hurry.

2.6 Additional Applications of Pick Detection

In this Section, I describe how our central idea and proposed approach of leveraging inertial sensors from a smartphone and a smartwatch to identify shopping gestures and individual item-level interactions is utilized in other extended applications of *shopping*. More specifically, I present the key ideas and my chief contributions in two other works [98, 107].

2.6.1 Key Ideas & Contributions

As an extension to *IRIS*, we first proposed an architecture [98] of combining sensor data from personal wearable-devices and store-deployed IoT sensors (e.g., BLE beacons) to infer item-interactions and the exact item being picked by a shopper.

Using various micro-studies on how users interact with objects placed on shelves in our lab (to crudely mimic similar in-store interactions), we establish two promising principles: (a) to better identify item-level interactions, we must utilize correlation between infrastructure and wearable sensor data; and (b) the camera on a wrist-worn smartwatch can identify a specific product selected by a shopper, but must be intelligently triggered to conserve energy.

We further proposed a solution, called I^4S [107], that also combines multiple low-energy BLE beacons, mounted on store shelves, with smartphone and smartwatch sensing to further identify the rack and shelf-level locations from where users pick specific items. The advantages of this approach is that in addition to identifying gestural interactions, it can also identify all the items that a shopper interacts with during the shopping episode. Such solutions are useful for both the shop owner (in capturing shopper's browsing interests) and the shopper (in obtaining real time individualized services).

Both the above ideas rely primarily on accurate identification of "item pick" gestures followed by fine-grained localization of such pickup gestures. This is achieved based on shopping gesture detection techniques proposed in *IRIS* based on fusion of mobile and smartwatch sensors (discussed earlier in Section 2.5.2).

2.6.2 Key Results

Our preliminary studies conducted in a lab setting achieves a 94% accuracy in identifying an item-picking gesture. Subsequently in I^4S , based on evaluation conducted with data obtained from 31 shopping visits at a mid-sized stationary store, we show that our pick detection approach achieves an accuracy of 92.85% (with a precision and recall of 92% and 81.5% respectively) using 10-fold cross validation and an accuracy of 89.18% (with precision of over 88%) when following a person-independent approach.

These results are based on studies conducted at different settings and with different set of people, and helps to validate the efficacy of our approach in accurately identifying shopping gestures.

2.7 Discussion Points

This chapter presents the design and initial prototype of *IRIS*, a framework for obtaining behavioral insights about a shopper's in-store interactions and behavior, utilizing only sensing data available from the shopper's personal smartphone and wearable device (smartwatch). Results show that, given a trace of an entire shopping episode in representative retail stores, *IRIS* is able to (i) delineate the (start, end) times of different in-store interactions, and (ii) utilize various shopping-related features to characterize such individual in-store interactions – both with very high (approx. 90%) accuracy. Overall, *IRIS* operates without any assumption of in-store infrastructure support or location tracking capability (no Wi-Fi, no RFID, no knowledge of store layout, etc.) and helps to build individualized *shopper-profiles*. There are a variety of additional approaches and possibilities in extending this work:

Inferring Multi-arm Shopper Interaction Gestures: In our studies, we have assumed that the shopper performs all the shopping gestures using their dominant hand. However, in real shopping scenarios, there will be cases when the shoppers might be using the alternate arm to pick or put-back the item. In such cases for example, if the shopper is wearing a smartwatch on one wrist and a fitness band on the other hand, the proposed *IRIS* approach would still be able to capture fine-grained shopping gestures. Given that it is not that common yet for people to use such multiple wearable devices, *IRIS* approach may need to be augmented with alternate infrastructural sensors for such finer-grained multi-arm gesture differentiations.

As we are primarily targeting those 'low-end stores' that do not have budget and techniques for capital-intensive investments, the preferred technology should be of low-cost. A plausible approach is to utilize the *short-range radar* devices which are becoming cheaper and are likely to become a part of future WiFi installations (e.g., as WiFi moves to 24 GHz, 60 GHz band). Such radar devices can be deployed in the store shelves/racks to track gestures performed by shoppers in front of the shelves. These recently emerging short-range radar devices [3] are capable of distinguishing between multiple moving objects in front and maybe able to capture shopping gestures irrespective of the limb(s) used. An added advantage of using such radar devices is that it can help capture shopping activities even when the shopper is not wearing a wearable device and also in cases where multiple users are engaged in the shopping activities (e.g., two people shopping together where one person pushes the trolley and other person picks individual items). Another possible way forward would be to use BLE beacons as a low-cost technology with minimal investment to attain additional capabilities.

Handling Scenarios of Multiple Individuals Shopping Together: IRIS is designed with the assumption that only a single shopper is visiting the store and performing all the shopping activities. However, there may be scenarios where multiple individuals visit the store, use the same shopping cart and shop together. In such a case, multiple shoppers may then 'inspect' the product, with one of them doing the 'pick' and another one doing the 'put in cart' action *or* one individual may simply push the cart while another does the picks. While in its current state *IRIS* would fail to accurately capture detailed insights in such scenarios, we believe that the system can be adapted to fuse data from multiple shoppers to properly capture the distributed shopping actions across individuals. The system should first identify the 'group' shopping behavior (e.g., identify 2 individuals shopping together) and then combine data sensed from personal devices of both shoppers to capture all item-level interactions. Based on some similarity in the locomotion patterns (e.g., step rate) and the hand movements involved, we may be able to identify the individuals who are shopping together.

Another assumption taken in *IRIS* is that shopper is using a trolley while shopping. We suppose that with minor fine tuning of certain functionalities, our approach would still work in situations where the shopper is carrying a shopping basket instead of a trolley. The key changes to be made will be in the strategy of differentiating between aisle vs non-aisle areas. While in the case when a shopper is pushing the trolley, there is minimal hand movement involved and thus, primarily leveraging the difference in locomotion pattern (e.g., inter-step interval) would help in identifying aisle vs non-aisle zones. On the other hand, when the shopper is carrying a shopping basket in hand, *IRIS* would need to distinguish between the hand movements when walking with a basket in a non-aisle zone and also the actions of 'putting down basket' and 'carrying the basket' within aisle zones. We believe that by combining such additional gestural insights and locomotion patterns, *IRIS* could adapt to scenarios where the shopper is using a shopping basket. However, when the shopper is holding the basket in the hand without a smartwatch, it would fail to capture certain insights and may not be able to achieve similar performance. Using data captured from multiple wearable devices worn on both arms of the shopper may then help in improving the performance.

Incorporating Physiological Sensor Data: Physiological sensor data (e.g., smartwatches contain embedded heart rate or GSR sensors) can help to additionally infer (or even *predict*) a shopper's in-store browsing intent and product-specific reactions. As a preliminary effort, we observed that using the mean and variance of heart rate values (captured by our smartwatch) allowed us to obtain a classification accuracy of 78% for item-level interactions (such as whether the shopper was picking a familiar item vs unfamiliar item). This also indicates the potential benefit in combining inertial sensing data with physiological sensor data to obtain more detailed inferences on user behavior during shopping activity.

Alternative Methods and Extensions: Given the potential for future applications in utilizing the insights on shopper's behavior inside a retail store, researchers have continued to propose alternative solutions. Here, I discuss some such key extensions for fine-grained in-store shopper monitoring that were proposed either at the same time or after *IRIS*. Shangguan et al. [110] proposed the *ShopMiner* system which uses RFID tags attached to individual items (in a clothing store) and exploit the backscatter signals of passive RFID tags to identify if the shopper is $\{$*looking at, picking up or turning over*$\}$ an item and also understand the relative attention they pay to different items. Unlike *IRIS*, this system fails to build an individual-level profile, which is an important aspect for personalized shopping applications. However, *IRIS* could be augmented by an approach like *ShopMiner* to identify the exact items the shoppers interacted with. We exploit the similar idea in building the I^4S system (explained earlier in Section 2.6). More recently, Zhang et al. [139] proposed the *ShopEye* system which identifies three kinds of relations (such as the user-item, user-user and item-item) in physical stores. Similar to the idea mentioned above, they utilize a hybrid RFID and smartwatch-based approach to delve into these relations and capture user behaviors and the item motions. There are also commercial solutions like Amazon Go [1] which provides a checkout-free shopping experience to customers, based on combination of video sensing and sensor (e.g., RFID tags) fusion technologies.

Plausible Additional Factors Affecting Shopper's Behavior: From our observations based on real-world user studies conducted at two different grocery stores, we perceive that the overall store layout and arrangements may potentially modify some of the behavioral assumptions. For example, the shopper's locomotion pattern may vary depending on the size of the store or the area of the aisles and non-aisles. Additional studies are required to confirm the impact of certain aspects such as wide vs. narrow aisles or whether stores run promotions (e.g., someone running a cooking session in a non-aisle area) on the system-level assumptions taken. In the long-run such factors may also be inferred using the *IRIS* platform to enable applications such as *crowdsourced store-profiling*.

The increasing trend for online or multi-channel search and consumption may also affect the overall shopping behavior of people–e.g., people may increasingly perform product research online and just come to the physical store to purchase specific items. We anticipate that such aspects may affect the shopper's 'browsing behavior' and in turn impact some of our item-level interaction features (e.g., the time taken to pick an item, overall time spent in specific aisles). However, we believe that the classifier models of *IRIS* could be evolved to incorporate these additional factors based on longitudinal observational data of individual shoppers across multiple shopping episodes.

2.8 Experiences and Lessons Learned

As part of this research, I initially observed individuals' shopping behavior in multiple retail grocery stores and conducted several experiments with real users in these stores. Here I outline some of the key learning points from this work.

- *Need to identify other gestures (e.g., inspecting an item)*: While this work focuses on identifying three main shopping gestures (pick, put-back, put-in-trolley), there are additional gestures that are performed by shoppers in a store. For example, a shopper may pick an item and "inspect" the item for some time and put it back. Similarly, in certain stores shoppers may "try out" specific items (e.g., trying out clothes, sunglasses). Identifying variety of such gestures would help in obtaining additional inferences like the time spent in inspecting or interacting with items (which may be of interest to the shopper).
- *Diversity of gestural interactions*: In practical shopping scenarios, the way certain shopping gestures are performed may vary depending on the object placement and/or the object size. For example, to pick up an item from the bottom of a shelf, the shopper may bent down first and then pick the item or when lifting heavier objects, the shopper may use both the hands together to get the item. The current gesture recognition model need to be augmented with sufficient training data (of such instances) to accurately identify similar gestures under varying conditions.
- *Need for multiple wearables*: While in our user study, the subjects were asked to wear the smartwatch on their dominant hand and perform the shopping

gestures with that hand, there were certain instances when some of them were using the other hand for picking up items. Similarly, as mentioned in the previous point there would be scenarios where shoppers would use both their hands to pick the item. Tackling situations like this would demand for having multiple wrist-worn wearables on the shoppers and requiring further mechanisms to use the sensor data from appropriate wearable or combining data from both.

2.9 Acknowledgments

This work was done in collaboration with Xerox Research Centre, India (XRCI). I extend my wholehearted gratitude to Dr. Sharanya Eswaran with whom I closely worked on this project. In Table 2.9, I summarize how the work was split between myself and Sharanya.

	Meera	**Sharanya**
Problem scope formalization & Approach planning	50%	50%
Data Collection	80%	20%
Implementation	90%	10% Survival Analysis in Section 2.5.3
Analysis	100%	

Table 2.9: Work and contributions split.

Chapter 3

Gym Monitoring and Digital Interventions

In this Chapter, I introduce an approach that utilizes only data sensed from a simple, cheap sensor device attached to the weight stack of an exercise machine for capturing fine-grained insights of individual's gym exercise behavior. To motivate this application and the chosen approach, I first present results of analysis of both digital gym usage records of 6513 individuals over a longitudinal period as well as survey of 575 gym-goers in Section 3.1.

3.1 Motivation for Wearable-Free Digital Tracking of Gym Exercises

Regular physical activity is essential to maintain good health, well-being and to stay fit [23]. As individuals become more aware of the benefits of engaging in physical activity, the prevalence of people going to the gym or fitness centers is on the rise. Recent statistics [115] report that the number of fitness center memberships in the United States has steadily increased over the last decade (with the membership count reaching 60.87 million in 2017). However, a major challenge among gym-goers seems to be the longer term adherence to the exercise behavior.

Prior studies in the behavioral literature [125, 132] have reported that participation in physical activity is influenced by a diverse range of personal, social, and environmental factors. However, little is known about the severity of the dropout problem, the temporal patterns exhibited by people who *dropout* (i.e., cease visiting the gym), and what other contextual factors seem to affect such individual-level dropout behavior.

Additionally, the rapid growth in the market for fitness devices and apps offers the possibility of providing quantified insights into an individual's exercise routine and enabling personalized interventions. Although there has been an explosion of such mobile applications for promoting healthful behaviors, relatively few have applied behavioral theory and lack aspects to get wider sustained adoption [57]. A review of such physical activity apps found that only 2% provided evidence-based guidelines for gym exercises training and report that these apps follow a one-size-fits-all approach and people find the recommendations or suggestions provided to be not helpful [59].

Given these facts, we believe that identifying key enablers for sustained gym participation, understanding what forms of failures in gym participation exist and what people desire to overcome such participation is important. Thus, we focus on studying the exercise habits of people, their temporal consistency or chances of dropping out and their reasons for quitting gym activity based on two kinds of data sources: (a) gym visitation data logs of 6513 individuals (captured through card transaction logs) visiting our University campus gym for a longitudinal period of 16 months and (b) survey of varying demographics of 575 individuals (of which 368 of them are a subset of the 6513 individuals for whom we had the gym visitation data from our campus gym) who are gym-goers or have stopped going. We also obtain insights on the desired features and services that people would like to have in a gym–these insights help us identify possible digital monitoring and intervention capabilities that may prove more effective in ensuring sustained participation in gym activities.

Overall, we believe insights presented in this work on gym user behavior draws attention to the need for improving the gym experience of people (in maintaining sustained participation) and helps to identify desired features for future digital intervention tools and motivates our proposed solution (discussed next in Section 3.2) for unobtrusive and personalized tracking of gym exercises.

3.1.1 Gym Visitation and Survey Dataset

In this work, our broader goal is to first obtain an overall understanding of gym usage behavior of individuals, investigate the retention and dropout rates of gym-goers and identify preferences of people in futuristic digital technologies for gym exercise tracking. To investigate these factors, we utilize two kinds of data–(a) the gym visitation data of users and (b) survey responses gathered from gym-goers about their gym usage behavior. Below we describe in detail both the datasets obtained.

3.1.1.1 Gym Visitation Data

We obtained the gym visitation data of users (recorded based on the tap in and tap out of user ID card at the gym) of our University campus gym for a continuous period of 16 months from September 2016 to December 2017 (including two fall terms, one spring term and one summer term). The gym tap-in/tap-out data log contains details such as the user ID, time of entry and exit for each visit to the gym and other demographics information such as gender, school of study, user type (e.g., undergraduate, postgraduate, exchange student, admin staff, faculty, alumni), year of study and course code (for students). After initial pre-processing and discarding of incomplete entries, the dataset we used included 94,188 data records from 6513 unique users who visited the gym during this period. We utilize this dataset to obtain aggregate usage statistics such as the temporal variation of gym usage pattern across a day/week/term and the dropout pattern of users, as well as infer factors that may help promote sustained gym participation.

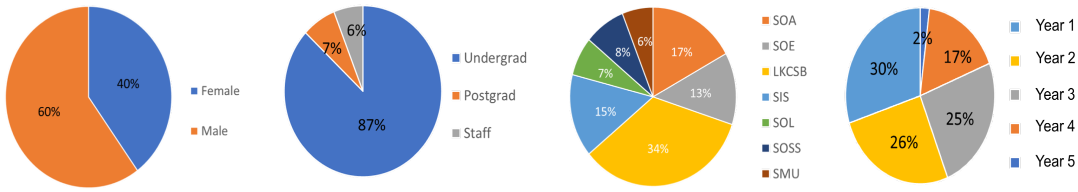
3.1.1.2 Gym Survey Data

We next conducted a survey to understand the gym usage behavior of individuals (e.g., reasons for going to or dropping out from gym, self-rated usage of specific workout zones or equipment in the gym), preferences or services that would help improve the gym experience of individuals, usage of fitness apps and key features desired from such digital tools etc. The survey was hosted in Qualtrics and was approved by our Institutional Review Board (IRB-18-028-A018(218)). This survey was conducted in two phases:

- *Survey distributed at University gym:* Distributed to the students and staff who visited the campus gym at least once during the academic semester for which we obtained the gym visitation data
- *Survey distributed to the Public:* Distributed to the general public members who goes to a gym.

Both the surveys consisted of 18 common questions (including 15 multiple choice and 3 open-ended ones). The survey distributed to the general public involved few additional questions (explained later in Section 3.1.1.2.2). The survey was designed such that the users rated the importance of specific statements under each question in a 5 point likert scale ranging from *"Not at all important"* to *"Extremely important"*. The survey also gathered other information from the respondents such as the frequency of their gym visits, the duration since the user has been going to a gym, self-rated usage of specific workout zones and exercise equipment in the gym, fitness apps used and reasons for liking or disliking those apps.

3.1.1.2.1 Survey at University Gym: The survey was distributed to 1960 users who are either students or staff in our University campus via email. We utilized the gym visitation data for one academic term to identify and send the survey to only those users who visited the campus gym at least once during this term. We obtained



(a) Gender (b) User Category (c) School of Study (d) Year of Study

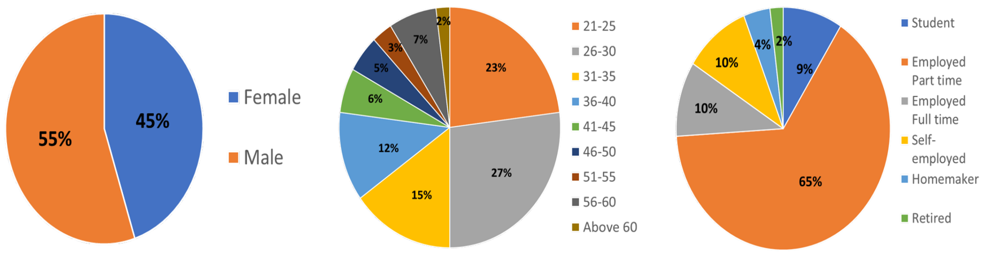
Figure 3.1: Demographics of University Gym Survey Participants

responses from 402 users out of which 34 were partial responses. A monetary compensation of 5$ was offered to the first 250 respondents.

In this survey, the respondents were categorized into three groups based on whether they (i) visited the gym at University campus, (ii) visited another gym or (iii) used to go to gym and has dropped out (i.e., ceases to continue gym activity after 1 or 2 visits). We ensured validity of these responses (at least for group (i) and (iii)) by comparing against the gym tap-in/tap-out data. While most of the questions were common to all groups, certain questions designed were targeted at specific groups.

We only collected the user email id in this survey. Further demographics information of the respondents are obtained from the gym tap-in/tap-out data mapped based on the respondent's email ids. In Figure 3.1, we report the basic demographic details of these respondents. Out of the respondents, 220 were males and 148 were females. 87% of the survey takers were undergraduate students. The highest number of responses were from the *School of Business* followed by *School of Accountancy* and *School of Information Systems*, which also corresponds to the school size. More than half of the survey respondents were freshers and sophomores, who also comprise the highest percentage of regular visitors at the campus gym.

We present results only based on full responses from 368 respondents. Among these respondents, 280 of them are regular visitors at our campus gym, 52 of them used to go to gym and has stopped going now and remaining 36 users goes to public gyms. Admittedly, this data has a strong demographic bias, as 87% of users are undergrads and thus likely to be *millenials*.



(a) Gender (b) Age Category (c) Employment Status

Figure 3.2: Demographics of Public Gym Survey Participants

3.1.1.2.2 Survey distributed to General Public: This survey was distributed online and was taken by the members of the general public in Singapore. In total, we obtained 207 responses, out of which 107 responses were obtained by distributing the survey to users of a community gym and the remaining 100 responses were obtained by hosting the survey in Amazon Mechanical Turk (AMT) (with respondent's location restricted to Singapore). The questions in this survey were similar to that of the one distributed in the University campus. As we lacked records of any actual gym visits or electronically captured profiles for these respondents, we also asked basic demographics questions such as age, gender, employment status. In this survey, we also included additional questions on the possible futuristic digital technologies (that would help provide a better gym experience and quantified tracking of workout activities to the individuals) and individual preferences and desired features for such digital tools.

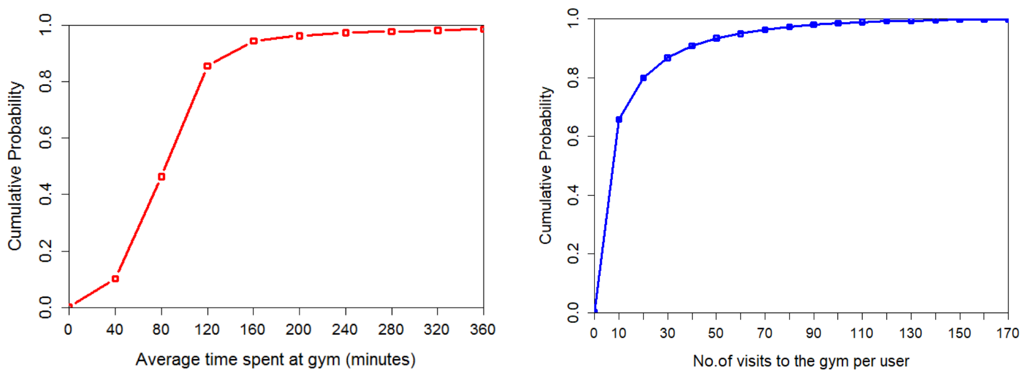
Out of these 207 respondents, 45 of them reported that they used to go to a gym and has stopped going now. The basic demographics details of these respondents are as reported in Figure 3.2.

3.1.2 Behavioral Patterns from Gym Visitation Data

We first seek to get a detailed understanding of the visit patterns and behavior of gym-goers using the *University* gym visitation data (explained earlier in Section 3.1.1.1). More specifically, we intend to study the following questions:

1. Do people exhibit regular visit patterns to the gym and does the gym visitation logs help uncover any temporal patterns in how individuals discontinue their gym visits?

2. Are there any key contextual factors that seem to affect the likelihood of continuing to visit the gym vs. dropping out?



(a)CDF of duration of continuous gym episodes across users

(b) CDF of total no. of visits to the gym per user

Figure 3.3: Cumulative distribution of the average time spend by users in gym and frequency of visit count

To understand the percentage of users who regularly visit the gym as well as those who dropout after one or two visits, we computed the total gym visit count per user for the period for which data was available. In our definition, *"dropouts"* constitute individuals who cease to continue their gym activity after less than or equal to two visits within 40 days of their first entry to the gym. We also refer to another category of individuals, *"infrequent visitors"* who visit the gym only few number of times (e.g., less than 10 visits over a 16 month period in our data) and the difference in days between their successive visits is high (greater than 40 days). The average time spent by 50% of the users at the gym is found to be about 80 minutes (see Figure 3.3(a)). A significant 15% of the users also spent more than 2 hours in the gym. Figure 3.3(b) plots the cumulative probability distribution of the total visit count of the users. We found out that over 65% of the users (i.e., 4283 out of the

6513 users) have less than or equal to 10 visits to the gym during the 16 months. More importantly, the rate of dropout (i.e., users with only 1 or 2 visits) was found to be 32%. This demonstrates that even in a gym where most of the gym-goers correspond to the student population, there is a significant set of users who dropout. Later in Section 3.1.3, we describe some of the key reasons why people discontinue their gym activity.

3.1.2.1 Acuteness of Dropout & Factors Affecting it

As discussed earlier, we observed that a significant percentage of gym-goers had only 10 visits or less to the gym (out of which 2071 individuals visited the gym only once or twice) during 16 months. For those individuals, we wanted to further investigate their dropout behavior–i.e., do most of them exhibit an early dropout behavior or are there individuals who also exhibit infrequent visit patterns? To study this, we first compute the average difference in days between an individual's successive visits to the gym and plot the cumulative distribution of it in Figure 3.4. This helps to distinguish between individuals who dropout from the gym after initial 1 or 2 visits and those who are infrequent visitors to the gym and still have only a 10 visits or less over a prolonged period. We found that 80% of the users dropout within the first month of visiting the gym and never return (i.e., their difference in number of days between successive gym visits were ≤ 40).

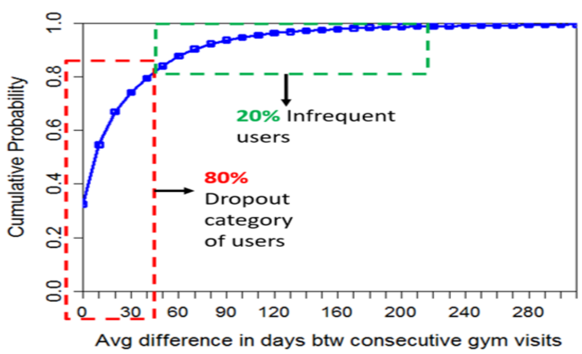


Figure 3.4: CDF of average difference in days between consecutive gym visits

Given that there is a high percentage of users stopping from visiting the gym, we were interested in understanding if there are any distinguishable behavioral patterns between regular gym-goers vs the dropout users. More specifically, we study two characteristics to see if they show noticeable differences between regulars and dropouts: (i) visiting the gym alone vs as a group (e.g., with a friend or an exercise group), (ii) regularity in terms of time of visit to the gym.

3.1.2.1.1 Difference in visit patterns–Groups vs Individuals: From the gym visitation data logs, we extracted the people who visited the gym as a group (i.e., with one or more individuals). For this, we first extracted all *user groups* whose gym entry time differences and exit time differences are both within *1 minute*–i.e., at an episode level, identify co-temporal gym visitors. We assume that people entering and exiting the gym within such short time gap visit the gym together and could be considered as in a group. Also, such joint visits should occur more than once to be declared as an actual group. As such, we extracted a total number of 1073 groups after discarding a count of 3416 singleton joint occurrences. Among these, 274 groups (25%) repeated five times or more. Also, 88% of these groups are 2 member groups and 10% are 3 member groups. This confirm that there is a trend of visiting gym as a group among users. Table 3.1 shows the breakdown of the repeated visit groups characterized by gender and school of study. We observe that 63% of the groups have members from the same school and 46% of them are *female-only* groups.

We next analyze the possible difference in visit patterns of individuals vs. those who come in groups. We obtained the cumulative distribution of the gym visit count for individuals vs groups (see Figure 3.5). The CDF plot shows that people going in groups visit the gym more number of times than people who go alone. Only 18% of the people who go alone has a visit count greater than 10 whereas for people visiting in groups it is greater than 45%. This indicate that visiting gym with a friend or as a group may increase the motivation to continue visiting the gym and thus minimize chances of dropout.

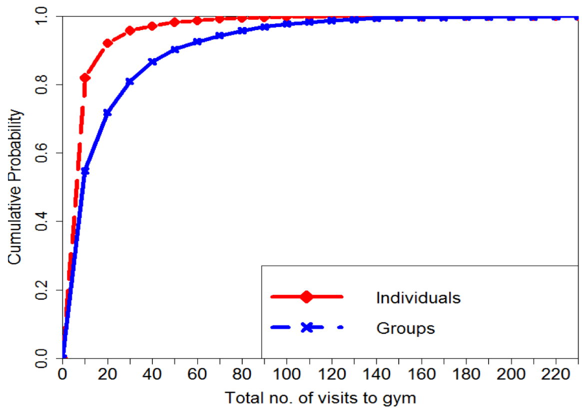


Figure 3.5: CDF of visit count for individuals vs groups

	All males	All females	Mixed	Same School	Different School
% in groups	28.8%	45.9%	25.3%	63.2%	34.7%

Table 3.1: Breakdown of people visiting gym in groups characterized by gender and school of study

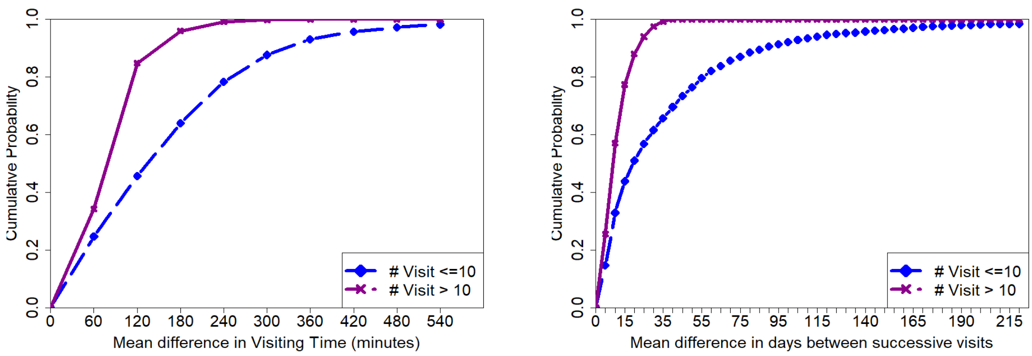
3.1.2.1.2 Regularity in visiting times–Regulars vs Dropout: We next examined the regularity in the visit pattern of individuals in terms of the time and days of visit and how it varied between those with a visit count greater than ten and less than or equal to 10 (i.e., regulars vs dropouts). i.e., Does individuals who continue to visit the gym regularly also exhibit a regularity in their visit schedule and are the time periods of visit more irregular for those dropping out? To investigate this, for each user we first computed the difference in their gym *entry time* and difference in the *number of days* between successive visits to the gym across all their records. This difference in visiting times is simply expressed as:

$$\Delta_t = EntryTime_{i+1} - EntryTime_i \tag{3.1}$$

Figure 3.6(a) plots the CDF of the mean of such differences in time of visit (i.e., mean of all $\Delta_t' s$ in minutes) for the two user categories ($\leq$ 10 visits and $>$ 10 visits). For those individuals with *visit count* $>$ 10, the difference in actual visiting times is within $\pm$2hours for nearly 85% of them (with 34% having a 1hr difference). However, for those with *visit count* $\leq$ 10, the Δ_t values were much higher (nearly

55% had $\Delta_t > 2$hrs), indicating greater irregularity in their actual time of visits to the gym.

We also computed the difference in number of days between successive visits and the exact days of visits for both category of users. We observed that people who visit the gym more number of times exhibit regularity in the days of visit to the gym (i.e., for example, an individual visiting the gym every two days or visiting only every Wednesdays). On the contrary, the individuals who had fewer visits barely exhibited any consistency in their visiting days or have longer gaps between successive visits. For example, from Figure 3.6(b) we can see that more than 40% of users with *visit count* $\leq$ 10 have a gap of more than 30 days between their successive visits (i.e., visited gym only once a month), whereas 78% of them with *visit count* $>$ 10 visited the gym at least once every two weeks.



(a) CDF of mean difference in visiting **times** between successive visits

(b) CDF of mean difference in number of **days** between successive visits

Figure 3.6: CDF of regularity in visiting time/days for those with *visit count* $>$ 10 and *visit count* $\leq$10

3.1.2.2 Key Takeaways:

- About 32% of people drop out or quit gym activity after 1 or 2 visits. Among these 80% of them completely stopped visiting the gym within their first month of visit.

- Going to gym in a group and following a regular gym schedule might reduce dropout and improve chances for sustained participation.

3.1.3 Insights from Survey on Gym User Behavior

Having obtained an understanding of the underlying behavior and visit patterns of individuals in a gym (characterizing a high rate of dropout), we next seek to primarily study the key reasons why people quit gym activity. We also intend to understand individual preferences and desired features that they would like to see in gyms for a better experience. From the survey responses gathered from 575 individuals (across different demographics), we aim to answer the following questions:

1. What are the key reasons why people discontinue and quit activity in a gym?
2. What are the desired features that people think would help in continuing their gym activity and improve their overall gym experience?
3. How valuable would it be for the users to have access to a personal trainer at the gym and what are the various things that a personal trainer could help them with?
4. What does individuals feel about the efficacy of existing fitness apps and wearable devices? Do they have any specific preferences in the technology they want to use while exercising in a gym?

Although the surveys were conducted in multiple phases, when presenting the results we combine the responses from all surveys, and highlight any differences in responses among different demographics, when applicable. Also, several of the questions in the survey were matrix table questions (i.e., ones that allow to ask and rate about multiple items in one question) with a 5-point likert scale rating. As such, when presenting the results, for each item in the multiple choice question, we combine the response count for the first two and last two scales (i.e., *"Extremely important"* & *"Very important"* and *"Not at all important"* & *"Slightly important"*) and ignore the neutral response (e.g., "Moderately Important").

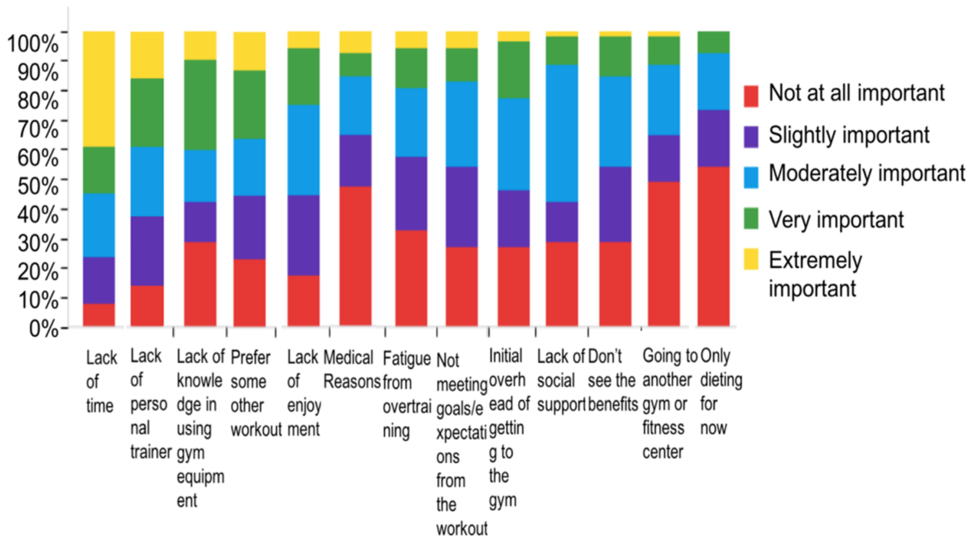


Figure 3.7: Survey response ratings on the reasons for quitting gym activity (x-axis labels sorted in descending order of importance)

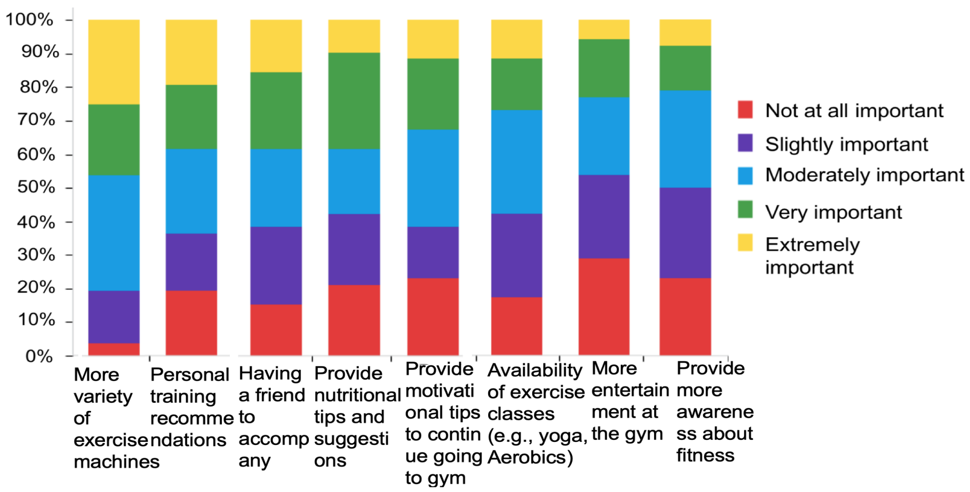


Figure 3.8: Survey response ratings for desired features/services to continue participation (x-axis labels sorted in descending order of importance).

3.1.3.1 Dropout Reasons & Desired Features to Continue Gym Participation

Out of the 575 survey respondents, 98 of them (17%) indicated that they used to go to gym and has stopped going now (or dropped out). In the survey, we specifically asked them the reasons for dropping out as well as the services that could help them to continue going to the gym.

As expected, *"lack of time"* is rated by 55% of the respondents as the main reason for quitting activity at the gym. More interestingly, *"lack of knowledge* *in using gym equipment"* (40.39%) and *"lack of personal trainer"* (38.43%) were among the top five reasons rated as important by the dropout users. This result holds across all the demographic groups (e.g., young, middle-aged, elderly) and suggests the two areas that could be improved to help the gym-goers. When asked about the services that would be important to the dropout users when deciding to continue going to the gym, the top response (46%) indicated a preference for "more variety of exercise machines". However, interestingly, *"providing personal training recommendations"* and *"having a friend to accompany"* were the next two common responses, rated as equally important by 39% of the users. The results of these two questions are as shown in Figure 3.7 and Figure 3.8 respectively. All the percentages reported are computed by combining the yellow and green blocks within each item in the x-axis of the plots.

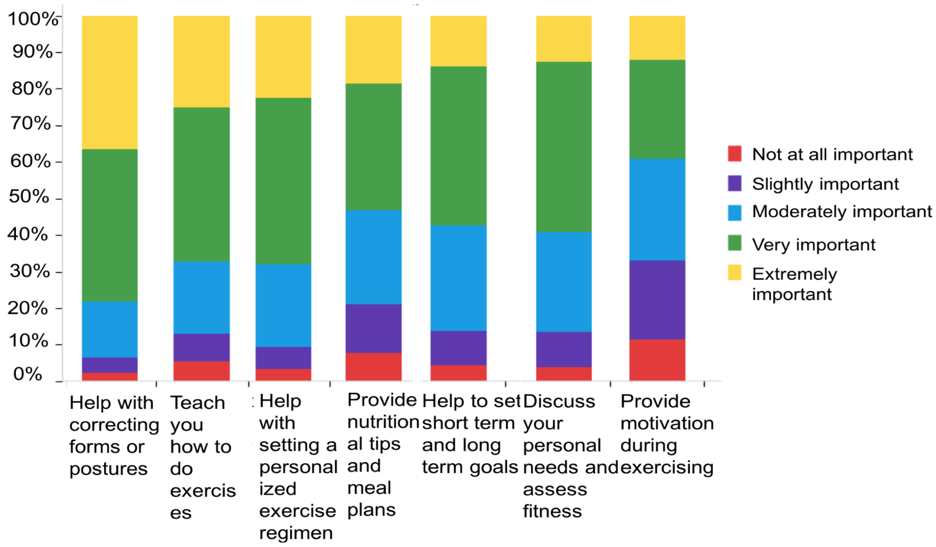


Figure 3.9: Services expected from a personal trainer (x-axis labels sorted in descending order of importance).

3.1.3.2 Need for a Personal Trainer

In the survey, we also included a question on the value of having access to a personal trainer in the gym and the key services that people would like to receive from a personal trainer. Having access to a personal trainer at the gym was rated as highly valuable by 44% of the respondents and another 22% of them rated it as moderately valuable. In Figure 3.9, we show the response ratings of the services that a personal trainer could provide. The survey responses also show that for 78% of the users across all demographics rate, *"help with correcting body forms/postures"* as the most important service a personal trainer could help them with. Other top-rated services from a personal trainer were to help with setting a personalized exercise regimen (68%) and to teach how to perform specific exercises (67%).

3.1.3.3 Usage of Fitness Apps

The next key question in the survey was to understand individual's affinity towards using a fitness application while exercising. To this question, 20% of the respondents stated that they are already using a fitness application, 63% expressed interest in using an app in the future and 17% responded that they stopped using fitness app(s). More importantly, over 70% of the people reported that they would be highly interested to use a fitness app that performs quantified exercise tracking and provides personalized feedback and corrective actions while exercising in a gym. People think that such recommendations would help make their exercise routine more effective and safer. As reported in the survey, some most common apps used by the individuals include *Apple Health*, *Samsung S Health*, *JEFIT* and *RunKeeper* and the commonly used wearable devices include *Apple Watch*, *FitBit* and *Garmin*. Individuals primarily used these fitness apps/devices to keep track of their cardio exercises, step count, heart rate and calories burnt. The people (97 out of 575) who discontinued using fitness apps reported the top reason to be apps not having met their expectations as the provided recommendations were too generic and not useful. Some of them also commented that using apps while exercising was a distraction from actual workout.

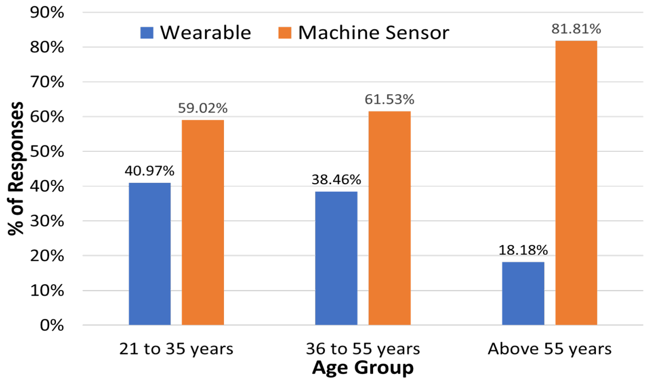


Figure 3.10: Preference of Wearable vs Machine Sensor-based technologies for people in different age groups

3.1.3.4 Adoption of Digital Technologies

The survey distributed to the public gym users also included a question on the preference of using a futuristic gym technology (which is either wearable OR a machine sensor-based technology) that can automatically track all the gym exercises performed and provide personalized quantified insights. From the 207 responses obtained, *59.9% (124 of those users) indicated an unwillingness to adopt wearable-based technology and preferred the machine sensor-based approach*. Notably, 82% of the users in the age group above 55 years were reluctant to adopt wearables, indicating a special adoption challenge among the elderly. This is also in accordance with the fact of digital aversion and the lower likelihood of using technology among elderly [30]. Figure 3.10 plots the preference for different age category of people. In general, the main reasons for the aversion towards wearable-based approach include: (i) the discomfort of wearing on-body devices and not wanting to use such devices while exercising, (ii) the inconvenience of requiring to wear multiple such devices for proper exercise tracking, (iii) forgetting to wear those devices and (iv) not wanting to spend money on wearables. Several of them who preferred the wearable approach over the machine sensor-based approach reported that they already own a wearable device and prefer it as it is more personalized and can also be used to track physical activities performed outside the gym.

3.1.3.5 Key Takeaways from the Survey:

The major takeaways from the survey are following:

- Top 5 dropout reasons– lack of time, lack of knowledge in using gym equipment, preferring some other workout, lack of personal trainer and lack of enjoyment.
- Providing personal training recommendations and having a friend to accompany are rated among the top services that could help the dropout users in getting back to the gym.
- 78% of the respondents reported that correcting form/posture is considered as the best service a personal trainer could help them with.
- 63% of the respondents are interested in using a fitness app and 20% are already using one.
- Nearly 60% of the individuals indicated a reluctance to use wearable devices while exercising, mainly due to the discomfort and intrusive nature of it.

3.2 Fine-grained, Practical Monitoring of Weight Stack-based Exercises

Given that retention or 'stickiness' for gym-based workouts remains a significant challenge based on prior studies [17] as well as our analysis of longitudinal gym data (described in the last section), there is a strong need for better mechanisms to support sustained gym participation. Moreover, an increased interest in gym regimens has also led to an increase in related injuries: between 1990-2007, over 970,000 people were treated in emergency rooms for weight training-related injuries, an increase of nearly 50% during the 18-year study period [55]. As such, solutions for automated, quantified and fine-grained tracking of gym activities (to

Figure 3.11: Common Weight Machines in Gym

maximize the workout effectiveness and reduce risk of injuries) are of high value in the fitness domain. The rapid growth in the market for IoT devices/sensors now offers the possibility of more quantified insights into a person's exercise routine, such as the type and intensity of exercise performed or the individual's *exercise form/posture*, with such insights enabling more personalized interventions such as exercise or corrective postural recommendations.

Most approaches for such quantitative capturing of an individual's workout activities rely primarily on either body-worn, wearable devices (e.g., [78, 141]) or infrastructure-driven video sensing [45]. Each approach has different potential drawbacks: (a) usability: wearable devices may not be readily adopted by the casual gym-going population (specifically, our survey (described earlier in Section 3.1.1.2) with 207 users in *community gyms* revealed that 60% were not in favor of using wearables), and a single wearable device may not be sufficient (e.g., wrist or arm-worn sensors cannot help track leg or hip exercises); or (b) privacy: video capture of workouts may be viewed as overly intrusive in public gym environments. Moreover, the efficacy of the techniques are typically evaluated over relatively short observational periods (e.g., 1-2 gym sessions).

We thus propose and evaluate a specific novel form of wearable-free and non-intrusive monitoring of gym exercises performed using weight stack-based machines. We hypothesized that such wearable-free monitoring might have some benefits–e.g., it might prove easier to deploy to some demographic segments. This justification is supported by empirical results provided in Section 3.1. Such weight stack machines (Figure 3.11) are widely used to perform activities for a variety of muscle groups. Our main intention to focus on the monitoring of exercises performed on "weight stack-based machines" is primarily due to the fact that other common exercise machines (e.g., treadmills, elliptical) already have monitoring that is in-built into the machines and moreover, there is an increase in weight training related injuries in the recent past [55]. It is worth reinforcing, at the outset, that the proposed approach requires *no user instrumentation* and utilizes only a simple, low-cost sensor device (with accelerometer and magnetic sensor) mounted on the weight stack (as illustrated in Figure 3.12), such that the sensor moves, dominantly along the vertical axis, during exercises. By applying an appropriate machine learning-based inferencing pipeline, we infer various exercise-related aspects simply from the *exercise motion-driven* variations in the sensor readings, in spite of the limited mode of observability (only vertical motion), noise and other user-specific artifacts. Our method extends prior work, on weight machines instrumented with multiple sensors (e.g., Jarvis [97]), with novel sensing pipelines to identify the user & the weight used and to accommodate medium time-scale changes in individual exercise patterns.

Given our minimalist approach (a single sensor, mounted at a single point) and the expected diversity in the range and type of exercises that different individuals perform, this work explores two fundamental **research questions**:

1. Can we build an inferencing pipeline, using data from only one simple weight-stack mounted sensor (which moves only vertically) to provide meaningful, multi-dimensional, fine-grained insights into the underlying exercise routine, such as 'amount of weight lifted' or 'which user is performing the exercise'? And, how does our accuracy compare with a wearable-based alter-

Figure 3.12: Multi-Purpose Cable Pulley Machine & Proposed Sensor Placement on the Weight Stack

native which directly tracks an individual's limb motions?

2. Is the inferencing pipeline, typically built through supervised learning based on labeled gym activity data collected over 1-2 sessions, robust enough to capture the medium-term *evolution* in an individual's gym activities? If not, how can the pipeline be modified, using incremental learning approaches, to ensure that it is able to *robustly* track the changes, over a span of months, in how an individual performs specific exercises?

Using a set of initial validation studies performed using a commonplace *multi-exercise* "cable pulley" weight machine, we develop a *multi-stage* pipeline (called *W8-Scope*[1]) to infer multiple novel facets of an exercise. We then conduct multiple larger-scale user studies *across 2 distinct gyms* (as described later in Section 3.4). Across these two gyms, we collected data from 50 participants performing 14 different exercise types with diverse weights, contributing 1728 sets of exercise data, over 103 distinct sessions, to validate the efficacy of W8-Scope approach.

3.3 W8-Scope : Overall Goals and Approach

W8-Scope's broader goal is to quantify various attributes related to exercising in

[1] pronounced Weight-Scope

a gym or a fitness facility. The gym-goers are interested in tracking their exercises, number of sets/repetitions, weight lifted etc. to understand their performance progress [62]. A review of physical activity apps found that only 2% provided evidence-based guidelines for resistance training [59]. Automatically logging the exercise performed, as well as the amount of weight lifted, helps users (especially novice or intermediate users who lack knowledge about the proper exercise posture or use of gym equipment) to track their exercise routine, progress and performance. A fitness monitoring application can integrate such monitoring based insights to provide personalized insights, such as : (a) Is the user committing more mistakes when performing *shoulder* exercises compared to exercises targeting other muscle groups? (b) Is the user training the same set of muscles repeatedly across different sessions? The application can then identify specific areas that a user needs to improve, or provide specific recommendations to help prevent serious injuries. In this work, we focus on the following facets:

1. identifying the amount of weight used
2. identifying the exercise performed
3. identifying possible incorrect patterns of performing the exercise
4. identifying which user is performing the exercise (the assumption being that each user has a unique signature while performing a specific exercise)

3.3.1 Design Goals and Challenges

Design Goals: One of our key goals is to devise a *wearable-free* and *non-intrusive* monitoring approach–i.e. infer the facets mentioned above without instrumenting the user's body with any wearable device. Our decision to avoid wearables is influenced not just on prior work [87] that suggests possible inconvenience from such devices, but also based on a survey that we conducted on gym-goers (explained earlier in Section 3.1.3). Also, for tracking the full range of typical gym exercises, wrist-worn devices are unable to track leg exercises, and one needs to adopt multiple wearables, placed on multiple limbs. Unlike previous approaches that have used infrastructural video sensing [45, 42] for exercise monitoring, we follow a *non-invasive* and *less privacy sensitive* approach. Specialized fitness facilities or gyms with advanced and expensive equipment (with built-in sensors) would have the capability of monitoring different exercise-related attributes. However, our goal is to also provide a *simple* and *cost-effective* solution. As such, we propose to use one or few simple small form-factor sensor devices mounted externally (i.e., after-market) on the top plate of a weight stack to infer the exercise and related attributes. Such an approach does not interfere with the normal usage of the exercise machine and enhances user convenience by not requiring the user to carry any on-body sensors. Unlike the Jarvis system [97] which also works using a machine-attached sensor, our approach of attaching the sensor on the weight stack of the machine helps to also identify the amount of weight that is lifted in addition to tracking other aspects of exercising. We evaluate our proposed approach on 7 different weight machines (including a multi-purpose "cable pulley" weight machine and six other machines that are dedicated for specific exercises).

Practical Challenges: Our proposed novel sensing mode using one measurement range for exercise monitoring poses the need for us to tackle several practical challenges: (i) Given that the sensor is attached to the weight stack of the cable pulley machine, distinguishing between different exercises becomes more challenging due to the movement of the weight stack, which can be similar across all exercises. This requires us to identify additional sensor-based features that could differentiate exercises; (ii) As the sensor is placed on the weight stack itself, it is thus exposed to noise, interference, and other confounding effects caused by nearby objects and users; (iii) Magnetic sensor is very sensitive to several environmental factors, including metallic equipment (e.g., dumbbells) carried by other gym-users; (iv) Different users perform the same exercise differently, based on the specific manner of execution, expertise in weight training exercises, physical strength and body build;

(v) Users exhibit inherent "drift" in exercising style across longitudinal time periods.

3.3.2 Overview of Final Design

We utilize a combination of accelerometer and magnetometer sensor streams from the *weight-stack attached* sensor to uncover various attributes of a set of weight-training exercises performed on the weight machine, while addressing the challenges described earlier. In our approach, to identify the amount of weight that is lifted, we mainly leverage the magnetic sensor data as the amount of magnetic field experienced by the sensor varies with different amounts of weight. We also combine features from accelerometer data to disambiguate magnetic sensor data which might look similar for different weights lifted to different heights. We then use a combination of features, extracted from both accelerometer and magnetic sensor, that is fed into a multi-stage classifier pipeline to identify the exercise performed[2], detect anomalous or incorrect exercise executions and also identify the user who is performing the exercise. Figure 3.13 shows an overview of *W8-Scope's* workflow.

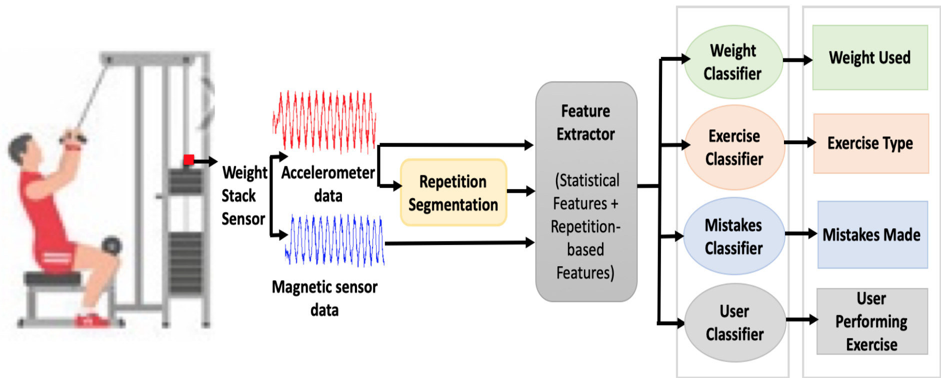


Figure 3.13: Overview of *W8-Scope's* Workflow.

[2]Note that such exercise differentiation is needed only for multi-purpose equipment.

3.4 Dataset

We conduct extensive studies and experiments with 50 users performing a variety of exercises on weight stack-based exercise machines under varying conditions. The data collection was performed in multiple phases at two different gym facilities (a *University* gym and a *Community* gym). The collected data included 3 distinct types of studies:

- *Controlled Study:* Conducted in the *University* gym, users were instructed to perform specific exercises with specific weights–these studies were used to identify relevant discriminative features and build the *W8-Scope* classification models.
- *Real-World Study:* Conducted at both the *University* and *Community* gyms, users exercised as per their own will–these studies, conducted in short-lived sessions, establish the real-world accuracy of *W8-Scope*.
- *Real-World, Longitudinal Study:* Conducted in the *University* gym, this effort involved a subset of real-world users being monitored across multiple weeks & months–the results helped in the design and evaluation of *W8-Scope's* incremental learning technique.

For the studies, we focus on a class of 14 exercises (listed in Table 3.2) that target different muscle groups and that the gym trainers indicated to be among the most popular exercise choices. At *University* gym, we monitored ten exercises performed using a weight stack-based "cable-pulley" multi-purpose equipment (shown in Figure 3.11). This machine has a set of 20 free-weights (each weighing $2.5kg$, except the top-most slab ($1.25kg$)), and permits at least 30 different weight training exercises [21]. Figure 3.14 shows the position of the exerciser and the weight stack during the upward motion of these ten exercises. In the *Community* gym, we utilize six dedicated single purpose weight machines for performing exercises such as *leg curls*, *leg press*, *triceps pushdown*, *biceps curls*, *chest press* and *shoulder press*.

Table 3.2: List of Exercises and their corresponding primary muscle groups targeted

Exercise Name	Primary Muscle Groups
Triceps Pushdown	Triceps
V-bar Pulldown	Lats
Biceps Curls	Biceps
Standing Cable Lifts	Abdominals
Bent Over Side Lateral	Shoulders
Seated Cable Rows	Middle Back
Single-arm Cable Crossover	Chest
Cable Rope Rear-Delt Rows	Rear-Delt/Shoulders
Upright Cable Row	Traps
Seated Two Arms Wrist Curl	Forearms
Shoulder Press	Shoulders
Chest Press	Chest
Leg Press	Quadriceps
Leg Curls	Hamstrings

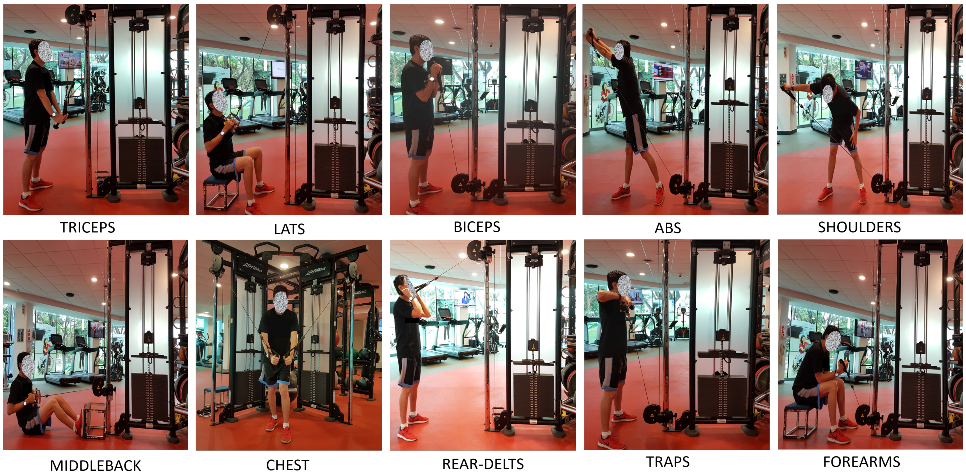


Figure 3.14: The exerciser's positions for first 10 exercises (performed on a multi-purpose cable pulley machine) listed in Table 3.2 in order.

These machines have varying number of weight slabs, weighing $7.5kg$ each. The users vary the cable heights and the amount of weights (adjustable using a pin) to perform various exercises.

We initially used a CC2650STK Sensortag device [121] developed by Texas Instruments (TI) to attach to the weight stack of the cable pulley equipment. However, after initial experimentation we replaced it with an iPhone 8 (at the *University* gym) as the accelerometer readings from the sensor tag turned out to be noisy and

Table 3.3: Summary of Data Collected and Results Obtained for the Controlled Studies

Experiment	Variations	Total no. of sets	Validation Results
[Different Weights] (with 3 exercises)	**9 weights** w=(3.75, 6.25, 8.75, 11.25, 13.75, 16.25, 18.75, 21.25, 23.75) kg for 3 exercises (biceps, triceps, lats); height as per user choice	54	**99.41%**
[Different Exercises] (with 2 weights)	**10 exercises** (biceps, shoulders, abs, traps, middleback, forearms, chest, rear-delts, triceps, lats) with 2 weights, w=(3.75, 6.25) kg	40	**98.74%**
[Different Heights] (with 1 exercise & 3 weights)	Lats exercise in which weight stack was lifted to **4 different heights** (6cm, 12cm, 18cm, 24cm) for 3 different weights,w= (3.75, 8.75, 13.75) kg	12	Mean error of $\pm 1.15cm$
[Different Sensor Positions] (with 2 exercises & 19 weights)	Lats and middleback exercise performed with sensor at **4 positions** (top and bottom center, top left and right corner) for weights varied from, w= $3.75kg$ to $48.75kg$	38	**98.96%** (with top & bottom sensors)
[Different Mistakes] (with 6 exercises & 1 fixed weight)	1 correct and **2 incorrect executions** (pull too fast, release too fast) for 6 exercises (abs, biceps, triceps, lats, chest, shoulders) with weight=3.75kg	108	**97.34%**

unreliable. We leverage the 3-axis accelerometer and 3-axis magnetometer sensors, sampled at a frequency of 50 Hz, from the iPhone 8 device. Note that, iPhone 8 is just used as a proof-of-concept and alternate sensors suffice. In fact, at the publicly accessible *Community* gym, where we could not leave the iPhone 8 attended, we have used an alternative multi-sensor device (DA14583 IoT Sensor[3]). We attach the sensor device to the top-most slab of the weight stack. This is a non-contact area of the user and did not affect the normal use of the equipment in any way.

3.4.1 Initial Validation Study

For the feasibility studies, we conducted a variety of experiments using the cable and pulley exercise equipment in our campus gym, over various controlled conditions across several days. The studies were conducted with multiple subjects including professionally trained gym staffs and other gym-goers. The key parameters that were varied in the study are: (i) the exercise performed, (ii) the amount of the weight lifted, (iii) the range of motion of the weight stack, (iv) different positions of placement of the sensor device, and (v) correctness of performing the exercise. In total, we collected 252 sets of exercise data (where a *set* is the number of cycles of *reps* completed; an exercise set in our study consisted of 10 reps, unless otherwise

[3]DA14583 IoT Sensor – (`https://www.dialog-semiconductor.com/iotsensor`)

specified) for different combinations of these parameters across 8 subjects (5 males, 3 females). Out of the 8 subjects, six of them are trained gym instructors and two are novices in weight training. All the exercise sessions were video recorded for ground truth purpose. Table 3.3 summarizes the experiments and the data collected as part of the controlled studies.

3.4.2 Real World Study

We performed user studies at our University gym and a Community gym in three different phases. The studies were approved by our Institutional Review Board (IRB-18-064-A052-M1(618) and IRB-18-153-A007(119)). For the user study at *University* gym, we recruited 35 (23 males, 12 females) university students and staff, who were in the age group of 21-35 years. For the study at the *Community* gym, 15 (9 males, 6 females) participants (age varying from 18 to above 60 years) were recruited. The subjects in both study included those with novice, intermediate and expert levels of expertise (self-rated) in resistance training.

3.4.2.1 Overall Study Procedure

Prior to data collection, each weight stack exercise machine was instrumented with a sensor, capturing both accelerometer and magnetometer sensors at 50Hz. The participants who agreed to take part in the study were required to visit the gym and perform a set of specified exercises. The participants were first briefed about the study and also shown videos of the exercises that they were required to do. At the *University* gym, the participants were also given a smartwatch (LG-Urbane), to be worn on their dominant hand, where a custom application captured accelerometer and magnetometer data (50Hz sampling frequency).

All exercises performed by participants were video recorded for obtaining the ground truth. From the experimenter's observation as well as based on the exercise videos collected, we found that the exercising style, pace, range of motion of

Table 3.4: Summary of real-world exercise dataset collected from *University* gym and *Community* gym.

	Study1_univ	**Study2_comm**
No. of participants	35 (23 males, 12 females)	15 (9 males, 6 females)
Age Variation	21–35 years	18–65 years
Self-rated expertise	13 (Novice); 16 (Intermediate); 6 (Expert)	9 (Novice); 3 (Intermediate); 3 (Expert)
No. of exercises	10 (targeted muscles: forearms, biceps, triceps, chest, abs, shoulders, rear-delts, lats, traps, middleback)	6 (targeted muscles: biceps, hamstrings, chest, quadriceps, shoulders, triceps)
No. of sets of exercises	Total 1148 sets of 10 reps each 320 sets (6 weights for 3 exercises from 18 subjects) 588 sets (10 exercises with 2 weights from 30 subjects) 240 sets (4 incorrectness for 2 exercises from 30 subjects)	Total 180 sets of 10 reps– 2 sets each of 6 exercises (with weights of subject's choice)
Variation of weights	6 weights ($3.75kg$ to $16.25kg$)	Weights used varied from $5kg$ to $80kg$
Incorrect exercise variations	4 (pulling too fast, releasing too fast, pulling half way through, lifting heavier weight)	N/A
Average duration of exercise session across subjects	48 minutes	19 minutes
Aggregated duration across all sessions	36 hours 50 minutes	5 hours 46 minutes

the weight stack, body posture varies across subjects. The number of sets and repetitions are as recommended by gym trainers and also as suggested in resistance training guide for healthy adults from the American College of Sports Medicine (ACSM) [83]. *Note:* In the user study, for every exercise set, we collected data for 10 repetitions each (unless otherwise specified). The participants were advised to take breaks (as required) in between exercise sets and were allowed to perform the exercises at a pace they are comfortable with. If subjects were not familiar with a certain exercise, it was first demonstrated to them by a gym trainer. Other than for the simulated incorrect executions, the subjects were not given any other special instructions and so, performed exercises *naturally*. An exercise session per subject ranged from about 35 to 55 minutes for $Study1\_univ$ and for 12 to 24 minutes for $Study2\_comm$. For participating in the study, we provided each participant a monetary compensation of $10.

3.4.2.2 Study in University Gym ($Study1\_univ$)

At our University campus gym, we conducted the user study with the multi-exercise cable pulley equipment in two phases. In the main study ($Study1\_univ$), we focused on collecting data for different exercises, different weights and simulated incorrect executions from 35 subjects. Among these, 30 participants performed:

(i) 2 sets each of the ten exercises (listed in Table 3.2), (ii) 3 sets of two exercises (*triceps* and *lats*) in a simulated manner such that they made mistakes such as “pulling too fast”, “releasing too fast” and “lifting only half through”, and (iii) 1 set of the same two exercises by “lifting heavier weights”. We ensured that all participants could easily simulate the mistakes in a safe and controlled manner by using a lighter weight of $3.75kg$. For lifting the heavier weight case, they were asked to choose a weight that they *perceived* as heavier than normal but within their comfort zone, separately for both *triceps* and *lats* exercise, and perform as many reps (up to a maximum of 10) that they could comfortably perform. This set of data was collected to mainly understand if users are more prone to committing mistakes (e.g., ‘releasing too fast’, ‘lifting only halfway’ and other ‘postural mistakes’) when lifting heavier weights. For obtaining data for different set of weights, 18 out of the 35 participants performed three exercises (namely, *triceps*, *biceps* and *lats* exercise) by varying it to 6 different weights (from $3.75kg$ to $16.25kg$). In total, we collected 1148 sets of exercise data. The details of this study are tabulated in column 2 of Table 3.4. For each of the 1148 sets of data obtained from the weight stack- attached sensor, we also obtained sensor data from a smartwatch worn by the participant while exercising. This data is obtained to compare the performance of our proposed *W8-Scope* approach to that of a more common and straightforward wearable-based solution.

3.4.2.3 Study in Community Gym ($Study2\_comm$)

As the university gym involved mostly student participants utilizing a single “multi-purpose” cable pulley machine, we utilized a community, publicly-accessible gym to obtain data from other demographic groups (e.g., working adults) and from different dedicated weight stack-based exercise machines. In this study (referred to as $Study2\_comm$), we collected data from 15 subjects (who widely varied in their age, employment status, ethnicity and expertise in weight training) performing 2 sets each of 6 different exercises (with weights of their choice) on the dedicated

Table 3.5: Summary of real-world *longitudinal* exercise dataset collected from *University* gym

	Study3_long
No. of participants	10 (7 males, 3 females)
Age Variation	21–35 years
Self-rated expertise	4 (Novice); 4 (Intermediate); 2 (Expert)
No. of exercises	5 (targeted muscles: triceps, biceps, abs, middleback, rear-delts)
No. of sets of exercises	Total 400 sets of 10 reps– 2 sets each of 5 exercises (with weights of subject's choice) on 4 different sessions
Variation of weights	Weights used varied from $3.75kg$ to $43.75kg$
Average duration of exercise session across subjects	14 minutes
Aggregated duration across all sessions	8 hours 20 minutes

weight stack machines. The targeted weight machines include the ones for performing *triceps pushdown*, *biceps curls*, *chest press*, *leg curls*, *leg press* and *shoulder press* exercises. In total, 180 sets of exercise data were recorded (see Table 3.5 for summary).

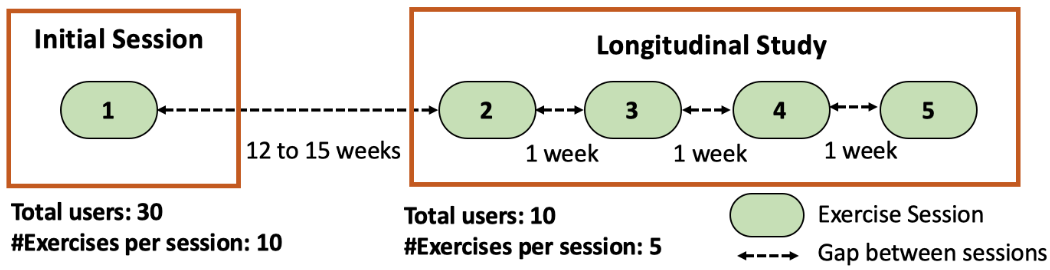


Figure 3.15: Longitudinal Study Period

3.4.2.4 Longitudinal Study in University Gym ($Study3\_long$)

In both $Study1\_univ$ and $Study2\_comm$, the users performed exercises in a single session. We further conducted a *multi-session* study ($Study3\_long$) with a subset of 10 users from the subject pool of $Study1\_univ$. In addition to the original ses-

Table 3.6: A high level summary of the main items discussed in Section 3.5 and the key takeaways from each.

Items Discussed	Key Takeaway
(Sections 3.5.1, 3.5.2): Accelerometer and magnetic sensor patterns for various exercises and the key observations supporting our system design.	Magnetic sensing, together with accelerometer-based height estimation, estimates 'weight lifted'.
(Sections 3.5.1.1, 3.5.1.2): Techniques to segment individual repetitions and compute novel features on it.	Novel features derived–displacement of the weight stack, repetition velocity, time taken to complete a repetition.
(Section 3.5.4): The multi-stage classification pipeline used in realizing the various facets of *W8-Scope* exercise monitoring approach.	Various facets derived–weight lifted, exercise type, mistakes made and user identification.
(Section 3.5.4.1): Results of different *W8-Scope* components based on the initial validation studies.	Accuracies for different components: Repetition count-98%, Weight lifted–99.4%, Exercise detection–98.7%, Mistakes identification–97.3%, user identification–99.1%

sion, these users performed exercises on 4 additional days (separated by a week); furthermore, there was a gap of over 3 months between the original session and these 4 sessions (Figure 3.15 illustrates the study period). In each of these session, the participant performed 5 exercises (namely, *triceps*, *biceps*, *abs*, *middleback* and *rear-delts*) with weights of their choice, resulting in a total of 400 sets of exercise data (details listed in column 4 of Table 3.4).

3.5 Design and Implementation of W8-Scope

To design *W8-Scope*, we first seek to get a detailed understanding of how the accelerometer and magnetic sensor data varies, as the weight stack moves while performing different exercises on a cable pulley machine. Table 3.6 summarizes the the key items discussed and the takeaways and insights that led to the final design of *W8-Scope*.

3.5.1 Accelerometer Sensor Analysis

We first inspected the accelerometer data recorded from the sensor attached to the free-weights stack while performing each of the first ten exercises mentioned in Table 3.2. We observed that the accelerometer z-axis data clearly shows the variation with each repetition and also varies across the 10 exercises, indicating the possibility of using an accelerometer to distinguish between exercises. The absolute value,

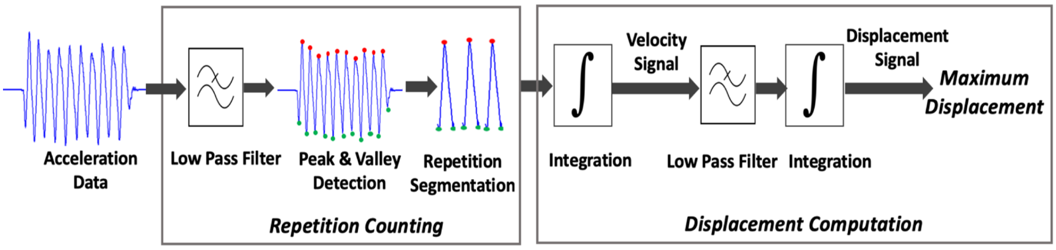


Figure 3.16: Steps involved in counting repetitions and computing displacement

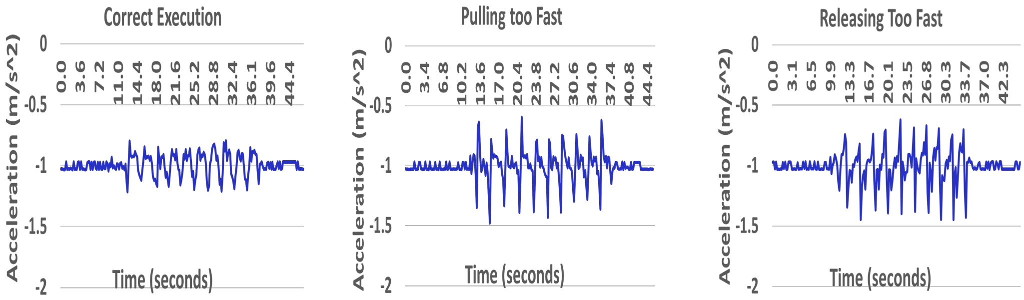
mean change and variation of the magnitude of acceleration and time taken per repetition varies across different exercises. Moreover, this also depends on the user and the pace at which an exercise is performed. In our controlled study data, the average time taken to complete one repetition across exercises is about 2 seconds (± 0.69 seconds standard deviation).

3.5.1.1 Identifying and Counting Repetitions:

To segment and count individual repetitions in an exercise set from accelerometer data, the following approach is taken (shown in *repetition counting* block in Figure 3.16). The raw accelerometer data is initially passed through a low pass filter. From the filtered acceleration data (for z-axes), we obtain the local maxima and local minima–i.e., points around which all other neighboring samples are lower/higher by δ (empirically set to 60% of the highest/lowest sample amplitude for our work). As certain repetitions were observed to have multiple peaks and valleys, an additional constraint on a minimum time threshold ΔT (empirically set to 2 secs) between successive peaks is used to avoid over counting. The time segment between two consecutive valleys is assumed to represent one repetition.

3.5.1.2 Computing the Range of Motion of Weight Stack

During our feasibility studies, we observed that one of the evident difference between exercises is in terms of the height to which the weight stack could be lifted (for the same amount of weights used). For example, when performing the *shoul-*



(a) Correct execution (b) Pull weights too fast (c) Release weights too fast

Figure 3.17: Variation in accelerometer readings while performing Triceps Push-down exercise (a) correctly & (b) by pulling weights too fast & (c) by releasing/slamming down the weights fast

ders exercise, the subject's arm has a wider range of motion (starting from a bottom lower level, the arms are stretched out at shoulder level), thereby lifting the weight stack to a higher level. The amount of time taken to complete a repetition also varies across different exercises. In addition, the weight stack's range of motion of the weight and the inter-repetition time vary for different amounts of weight lifted (e.g., lifting heavier weights would take longer time). To compute the weight stack displacement (outlined in Figure 3.16), we first extracted the z-axis acceleration signal, integrated it using cumulative trapezoidal integration[4] to obtain velocity, then low-pass filtered and then integrated again to obtain the displacement. As shown in Section 3.5.4.1, this approach results in a mean displacement error of 1.15 cm.

3.5.1.3 Understanding Quality of Exercise Repetitions

To understand the feasibility of identifying mistakes while exercising, we first consulted the professional trainers in our campus gym to understand the common mistakes that people make while training with the weight machines. As reported by the gym trainers, (a) pulling or releasing the weights too fast, or (b) lifting the weight only half way through corresponded to some "common mistakes" made by novice users. However, in certain workouts, explosive training techniques are employed to

[4]cumtrapz() function in pracma package of R – (`https://cran.r-project.org/web/packages/pracma/pracma.pdf`)

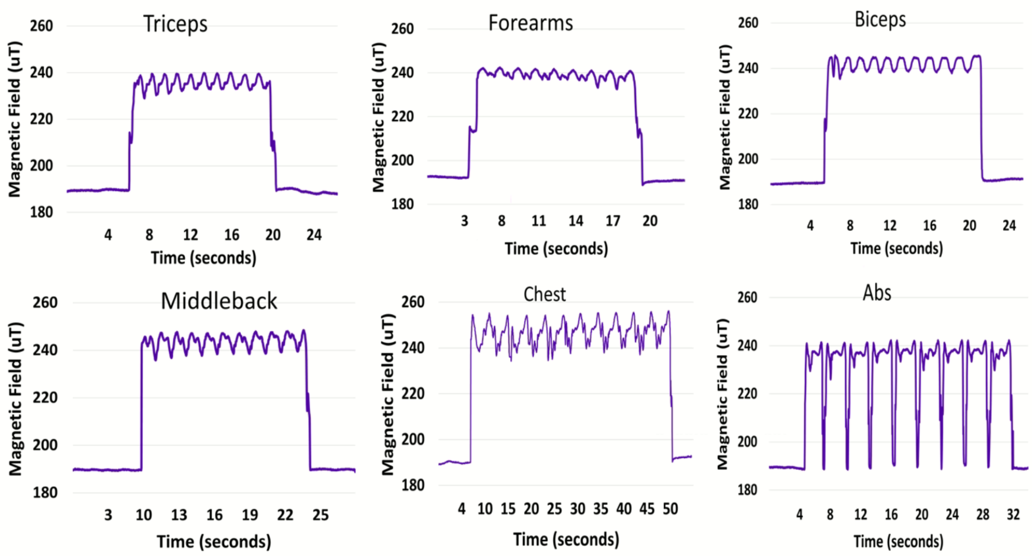
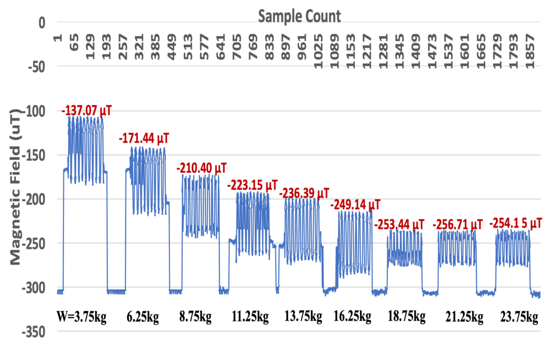


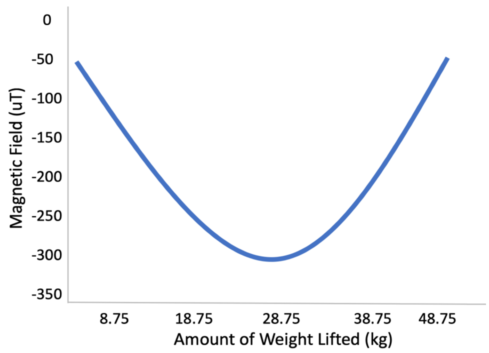
Figure 3.18: Variation in magnetic field for a sample of 6 exercises performed with weight, $w = 6.25kg$

improve or sustain muscle activation. In such cases, pulling the weight stack too fast or having only a smaller range of motion are legitimate means of power training. However, based on our interactions with multiple professional gym trainers, we understand that such techniques are used principally by expert athletes looking to target specific muscle groups. For our target group of novice/intermediate users, such patterns are almost always "mistakes" that can be corrected through explicit feedback.

As a preliminary study, we focused on collecting data for 6 different resistance training exercises (listed in Table 3.2) targeting *abs, biceps, triceps, lats, shoulders* and *chest* muscles. We collected data from 6 trainers at the gym for 3 sets of 10 *reps* of each exercise. Out of the 3 sets, they were instructed to perform one set correctly and two sets incorrectly–i.e., pull the weights too fast or release the weights too fast. We found (e.g., see Figure 3.17) that the accelerometer data contains visible *signatures*, that can help distinguish between such correct and incorrect execution patterns (as shown later in Section 3.5.4.1).



(a) Variation in magnetic field for 9 sets of 10 repetitions of lats exercise with weights per set varied between $w = 3.75kg$ and $w = 23.75kg$. (The number on top of each set shows the mean value (μT) of the magnetic field when lifting specific weight).



(b) Expected "U-shape" of the magnetic readings as the amount of weights is varied from minimum weight ($3.75kg$) to maximum weight ($48.75kg$).

Figure 3.19: Variation in magnetic field for different weights

3.5.2 Magnetic Sensor Analysis

We also studied how the magnetic field, sensed by a magnetometer, varies when performing different exercises using the cable pulley weight stack machine. We found that the magnetic field indeed varies as the weight stack goes up and down, indicating individual repetitions of each exercise. Figure 3.18 shows the distinct pattern of the magnetic field for the ten exercises performed with a weight, $w = 6.25kg$.

3.5.2.1 Variation in Magnetic Field for Different Amounts of Weight Lifted:

More interestingly, when we analyzed the data collected from the magnetometer for the same exercise performed with different set of weights, we observed that the magnetic field changes as a function of weight lifted. To understand this analytically, consider the weight stack has a set of m weight slabs, each slab with mass=w. Let d_i be the distance of the i^{th} slab from the sensor, while at rest, and let D be the distance (height) moved by the set of K ($K \leq M$ weight slabs that are lifted). In Equation (3.2), we represent magnetic field strength, B as a function of K. We can break this up into the part of the weight (the K slabs) that moves up (leaving the slab-sensor distance unchanged) and the part (the $M - K$ slabs) that don't move (leading to an increase in the slab-sensor distance).

$$B = \sum_{i=1}^{K} \frac{w_i}{d_i^2} + \sum_{i=K}^{M} \frac{w_i}{(D + d_i)^2} \tag{3.2}$$

Mathematically, the magnetic field of any body is inversely proportional to the square of the distance. Accordingly, as illustrated in Figure 3.19(b), the magnetic field at the zenith should exhibit a U-shape curve, initially decreasing (as k increases from a small value) but then eventually increasing (as the first term begins to dominate when k becomes larger), as a function of the weight lifted.

Figure 3.19(a) shows the variation in magnetic field while performing 10 repetitions each of *lats* exercise with 9 different set of weights ranging from $3.75kg$ to $23.75kg$. The figure is annotated (in red color) with the mean value of the magnetic field as experienced by the sensor when lifting varying amount of weights. Initially as the amount of weight is increased, the strength of the magnetic field keeps decreasing, thus making it easier to distinguish between the lighter weights. However, at higher weight values, the differentiation in the magnetic field is less pronounced (e.g., the mean magnetic field is -255μT when lifting either $w = 21.25kg$ or $w = 23.75kg$).

3.5.2.2 Magnetic Field for Different Weight Stack Heights:

Given that the magnetic sensor shows distinguishable trends for the different exercises and for different amounts of weight lifted, we wanted to answer the question: would just a magnetic sensor on the weight stack suffice or are there indeed cases where the magnetic sensor would be unable to distinguish between "weight=$w1$, height=$h1$" and "weight=$w2$, height=$h2$" combinations? We conducted an experiment in which *lats* exercise was performed with 3 different weights ($3.75kg$, $8.75kg$, $13.75kg$) lifted to 4 different controlled heights ($6cm$, $12cm$, $18cm$, $24cm$). We observed that the change in magnetic field for weight, $w = 8.75kg$ and height,

Table 3.7: Features extracted from each time window of accelerometer and magnetometer data

The *Count* represents the number of signal axes on which the feature is computed. For example, *count=4* means features are extracted on the *x-axis*, *y-axis*, *z-axis* and the *magnitude* of the signal and *count=1* means features are extracted on the *z-axis* of the accelerometer signal (as it showed clear variations pertaining to the exercise motion).

Feature	Count	Description
Mean	4	Average of the values for the time window for each axes and the Euclidean norm (magnitude) of the signal
Max	4	Maximum value in a time window for each axis and signal's magnitude
Min	4	Minimum value in a time window for each axis and signal's magnitude
Range	4	Total change in values within the time window for each axis and signal's magnitude
Variance	4	Variance of the values in a time window for each axis and magnitude of signal
Spectral Entropy	4	Normalized information entropy of the FFT components of each axis and magnitude of signal
Spectral Energy	4	Mean value of the square of the FFT coefficients of the signal for each axis and magnitude value
Mean crossing rate	4	Number of times the values cross the mean of the time window
Covariance	3	Covariance between each pair of axes of the sensor
Correlation	3	Correlation between each pair of axes of the sensor
Repetition Time	1	Average time taken to complete a repetition in a exercise set
Repetition Height	1	Average height to which the weight stack was lifted within a set
Repetition Velocity Mean	1	Average of the speed with which the weight stack was lifted in a set
Repetition Velocity Std.dev	1	Standard deviation of the speed with which the weight stack was lifted in a set

$h = 6cm$ looked very similar to that of $w = 13.75kg$ and $h = 24cm$ (mean and total changes being approx. 45μT and 32μT respectively for both cases). This shows that pure magnetic sensor alone can provide ambiguous results, and that both magnetic and accelerometer sensor data are thus needed to accurately distinguish between different weights lifted during different exercises.

3.5.3 Sensor Data Analysis: Key Takeaways, Data Processing and Features

Based on our controlled experiments and data analysis, our major takeaways are: (i) the weight stack movement is clearly identifiable from the magnetometer data, (ii) the accelerometer sensor can provide an accurate estimate of the precise exercise-related movements, with distinct $z - axis$ patterns for different exercises, (iii) using the accelerometer data, it is possible to derive two useful features: the *time taken* to complete a repetition as well as the *height* to which the weight stack is lifted, and (iv) the variation in the magnetometer readings can be used to identify the amount of weight that is being lifted (in conjunction with the displacement height estimate obtained from accelerometer data).

Accordingly, we use both the accelerometer and magnetic sensor data to identify various exercise-related attributes. Both streams of sensor data per individual exercise set are first pre-processed to remove any outliers. The pre-processed sensor data is divided into frames of length *w* ($w = 2$ seconds, based on the observed duration of a single repetition). On each frame of the signal, we first extract statistical features. The features were computed for each axis as well as magnitude of both accelerometer and magnetic sensors. Using methods described earlier in Section 3.5.1, we also compute *repetition-based* features such as average *time taken* to complete a repetition in an exercise set, average *height* to which to the weight stack was lifted during each exercise, and the average & standard deviation of *speed* with which the weight stack was lifted and brought down. The complete set of features used in our classifier models is listed in Table 3.7.

3.5.4 The W8-Scope Classification Pipeline

Based on the insights gathered from the sensor data analysis, we develop the *W8-Scope* classification pipeline that leverages on specific features that are extracted from the accelerometer and magnetometer sensor data. Using these sensor-based features, we first identify the amount of weight lifted and then identify the exercise performed. Subsequently, in logically-parallel steps, we detect incorrectness in specific exercise execution and distinguish between the users performing same exercise.

Initially, we tested performance of various classifiers (SVM, Decision Trees, Random Forest) in *Weka* [43] for classifying the different weights using the window-based features extracted from data collected during the controlled study for *different weights* experiment (explained in Section 3.4.1). We first tuned the parameters of the different models on our dataset and selected the parameters that gave the best

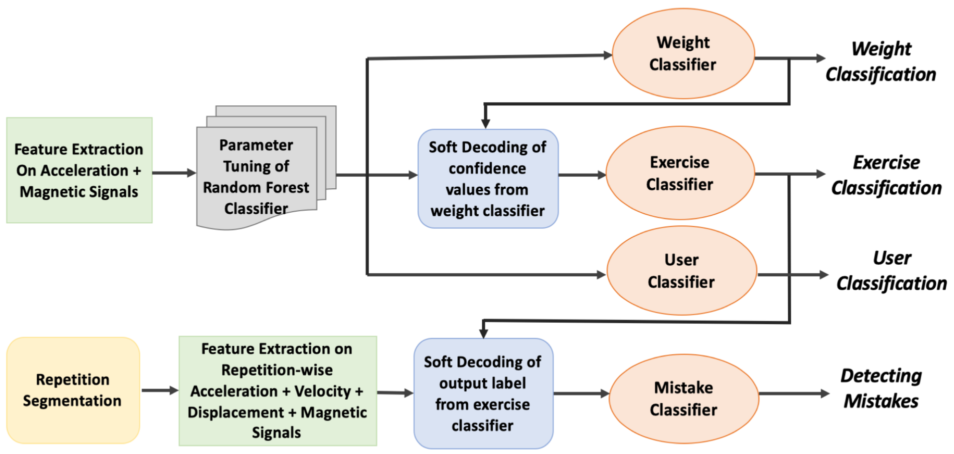


Figure 3.20: Pipeline of classifying the amount of weight lifted, exercise performed, user performing the exercise and incorrect exercise executions made

performance for each model. We then evaluated the performance of the parameter tuned machine learning models using 10-fold cross validation and found that the best classification performance was achieved with Random Forest (RF) Classifier (with *number of trees*= 60). Hence, we used RF classifier throughout our multi-stage pipeline–this is consistent with prior works (e.g., [46, 127]) that also found RF classifiers to be more accurate for sensor-based exercise monitoring.

The key components in the classification pipeline (see Figure 3.20) are as follows:

- *Amount of Weight Lifted Identification* – We train a *weight classifier* using the parameter tuned random forest classifier. The weight classifier provides the classification of the different weights and the distribution of confidence values for the set of weights (i.e., the probabilities that weight = $[w1, w2, w3, w4, w5, w6]$) for each instance.

- *Exercise Identification* – For the *exercise classifier*, we follow a *soft decoding* approach, that is to feed in the results from the prior classifier as a new feature vector to the existing set of features, i.e., we use the probability distribution of weights classification, instead of using only the 'most likely' label for the weights. The exercise classification is performed on the new feature set with the parameter tuned RF classifier.

- *Detecting Mistakes in Exercise Execution* – After identifying the exercise, we attempt to detect the mistakes made, at a *per-repetition* level. (This is necessary as users may incorrectly execute only a subset of the multiple repetitions in a set.) We first segment acceleration and magnetic sensor signals corresponding to the upward and downward motion of the weight stack during a repetition using techniques described earlier in Section 3.5.1.1. We also obtain the *velocity* and *displacement* corresponding to each transition. Instead of using a fixed window size, we now extract the statistical features on frames representing individual transitions for four signals (*acceleration, velocity, displacement and magnetic*). We also feed in the output of the *exercise classifier* as a new feature, by taking majority output labels during a set. Note, as shown in Figure 3.20, this implies that mistake classification is not real-time–i.e., it is only performed retrospectively, after the user has completed an entire set (usually lasting 30-40secs). On the new set of features extracted, we again used a RF classifier to classify the commonplace mistakes such as "pulling the weight stack too fast", "releasing fast or slamming down the weight stack", "lifting the weights only half-way through".

- *User Identification* – This component is used to distinguish between users performing the same exercise on the cable pulley weight machine. For this purpose, we used the initial set of features used for weight classification and split it into *exercise-specific* feature files, subsequently building a per-exercise classifier that attempts to predict the exercising user, given an entire exercise *set*.

3.5.4.1 Controlled Study Results

We now present summarized results of the different *W8-Scope* components, evaluated on controlled studies performed initially with a small set of explicitly-instructed users (explained earlier in Section 3.4.1). As these studies do not cap-

Table 3.8: Average error (in cm) in displacement computation for varying heights to which weight stack is lifted

Actual Height	6 cm	12 cm	18 cm	24 cm
Average Error	±0.67 cm	±0.87 cm	±1.1 cm	±1.96 cm

Table 3.9: Controlled Study –Summary of performance accuracy (using 10-fold cross validation) for each classifier using individual sensors as well as combination of both sensors

	Weight	**Exercise**	**Mistakes**	**User**
Only Accelerometer	77.49%	91.53%	90.43%	93.41%
Only Magnetometer	92.96%	79.37%	83.85%	87.65%
Accelerometer and Magnetometer	**99.41%**	**98.74%**	**97.34%**	**99.12%**

ture the natural gym activities (e.g., the weight variations across exercises, the sequence/mix of exercises performed), the results here are meant primarily to quantitatively differentiate the capabilities of the magnetic vs. accelerometer sensor, and to establish the accuracy of several of the key *W8-Scope* features (rather than the inferred outcomes).

Repetition Counting: Based on the 94 sets (containing 940 repetitions) of data collected from the different {*weights,exercise*} combinations, we ascertain that the repetition counting mechanism (Section 3.5.1.1) achieves an accuracy of **98%** in counting the 10 repetitions in each set.

Weight Stack Displacement: We studied the accuracy of displacement estimation (i.e., how much did the weight stack move during a repetition?), using the data collected from controlled *lats* exercises, where the participant lifted the weight stack to four different heights ($6cm$, $12cm$, $18cm$ and $24cm$) for three different weights ($3.75kg$, $8.75kg$ and $13.75kg$). We observed an average estimation error of $\pm\mathbf{1.15}\boldsymbol{cm}$ compared to the ground truth height. Table 3.8 shows the breakdown of the average error in displacement computed for each height.

Weight Amount: We utilized the data collected from 54 sets (from two subjects) for three exercises (*biceps*, *triceps* and *lats*), with weights varying from $\mathbf{3.75}$–$\mathbf{23.75}\boldsymbol{kg}$. The RF classifier achieves an accuracy of **99.41%** (yielding an average precision of 0.992 and recall of 0.994) in distinguishing between the 9 set of weights. In contrast, the accuracy for weight classification using *only magnetic* and *only accelerometer* sensors were 92.96% and 77.49% respectively, showing the importance of fusing multiple sensing modalities.

Exercise Detection: Using the data collected for 2 sets each of 10 different exercises, we found that using only accelerometer and only magnetic sensor based features result in an exercise classification accuracy of 91.53% and 79.37% respectively, whereas the joint use of features results in an overall performance accuracy of **98.74%** (with an average precision of 0.988 and recall of 0.987) in distinguishing between exercises.

Identifying Mistakes: We used the *W8-Scope* pipeline (Section 3.5.4) to perform a multi-class classification {correct, incorrect-pull fast, incorrect-release fast} on the data provided by 6 gym staff, which included deliberate mistakes in exercise execution. The performance accuracy achieved when using *only accelerometer*, *only magnetometer* and *combination* of both sensors were 90.43%, 83.85% and **97.34%** respectively.

Distinguishing Users: Using the data collected from 8 subjects (48 exercise sets), we found that *W8-Scope* can distinguish users (i.e., distinguish between the 8 users performing a specific exercise) with an accuracy of **99.12%** (precision of 0.991 and recall of 0.993) when using a combination of both sensor features, with the accuracy dropping to 93.41% and 87.65% when only accelerometer or magnetometer features are used.

Summary: Table 3.9 summarizes the key numerical insights. Our controlled studies show that *W8-Scope* can be promising (accuracy of over 97% using 10-fold cross validation) in realizing each of the attributes in *W8-Scope*, and that combining both accelerometer and magnetic sensor based features helps to increase system accuracy.

3.6 Real-World W8-Scope Evaluation

We now present the performance evaluation of *W8-Scope*, along with insights gained, based on real world, naturalistic exercise data collected from two gyms: (a) a *University* gym, that is equipped with a single multi-purpose machine and is primarily used by university students, and (b) a *Community* gym that contains multiple exercise-specific machines and is used by a wide variety of neighborhood residents. We focus on the primary attributes of interest {*Weight Used, Exercise Performed, User Identity, Mistake Identification*}, studying *W8-Scope* performance under real-world conditions. For the *University* gym, we also compare our proposed approach against that obtained via a wearable (smartwatch). Intuitively, a smartwatch should be able to more accurately distinguish between the different weight training exercises performed as the range of motion of the arm (including the orientation, starting and ending positions of arm) vary for different exercises. We also present additional behavioral insights obtained from manual annotation of exercise videos.

3.6.1 Counting Repetitions

The time taken to complete a repetition and also the displacement of the weight stack are used as features in our classification model. As computing these features require accurate segmentation and counting of individual repetitions in a set, we first evaluate the performance of *repetition counting*.

Using 908 sets of data collected from different weights and different exercises experiment in $\boldsymbol{Study1\_univ}$, we obtained a performance of $\mathbf{97\%}$ in accurately counting the 10 repetitions per set. Out of the 28 incorrectly counted sets (that caused 3% error in counting reps), 12 sets are off by $\pm\mathbf{1}$, 9 sets are off by $\pm\mathbf{2}$, 4 set are off by $\pm\mathbf{3}$, 2 sets are over counted by 4 and 1 set is under-counted by 5. *W8-Scope* under-counted the repetitions primarily for the *forearms* exercise, because the range of motion of the weight stack was too short to show evident peaks in acceleration data. Over counting of repetitions happened mainly when the subject moved

Table 3.10: Performance of identifying the amount of weight lifted – weight stack vs. wearable sensor

Weight Classification	**Accuracy**	**Precision**	**Recall**
10-fold CV using Weight Stack Sensor Data	97.5%	0.978	0.971
LOOCV using Weight Stack Sensor Data	93.75%	0.937	0.938
10-fold CV using Smartwatch Data	84.37%	0.822	0.845

the weight stack up and down in the beginning of the set while they were prepping to get started. For the 180 sets of additional data collected from $\boldsymbol{Study2\_comm}$, the repetitions were accurately counted for 177 sets (98% accuracy), indicating that this estimation was accurate across gym environments.

3.6.2 Amount of Weight Lifted Identification

We evaluate the performance of weight classification on different weights' data obtained from $\boldsymbol{Study1\_univ}$. Based on 10-fold cross validation with RF classifier (which outputs the dominant label observed across all the repetitions in a set), we achieved an accuracy of $\mathbf{97.5\%}$ in distinguishing between six set of weights, *w*=$[\mathbf{3.75, 6.25, 8.75, 11.25, 13.75, 16.25}]$ in the weight stack, with the classification error confined to the heavier weights – $\boldsymbol{13.75kg}$ and $\boldsymbol{16.25kg}$.

We also performed a *leave-one-subject-out cross validation* (LOOCV) in which the *weight-classification* model was trained with data from all users, except the test user, and then tested on the data from test user. Using this approach, we obtained an average accuracy of $\mathbf{93.75\%}$, with a precision of 0.937 and recall of 0.938 in classifying the weights, i.e., the mean percentage error was 6.25%, with the maximum error (11%) in recalling weight, w=$\boldsymbol{16.25kg}$.

We also evaluated the performance of *weight-classifier* on the smartwatch data. For this, we obtained an overall accuracy of 84.37% (precision=0.822 & recall=0.845) for classifying six different weights. Clearly, a weight-stack mounted sensor is able to identify the weight lifted more accurately than a hand-worn sensor. Table 3.10 presents the summary of results from weight classifier.

Predicted Label	Actual Label									
	abs	biceps	chest	forearms	lats	middleback	reardelts	shoulders	traps	triceps
abs	98.0%	0.2%	0.1%	0.1%	0.0%	0.2%	0.2%	0.7%	0.2%	0.0%
biceps	0.1%	96.1%	0.4%	0.3%	0.2%	0.5%	0.5%	0.3%	0.4%	0.2%
chest	0.2%	0.1%	96.5%	0.0%	0.3%	0.3%	0.0%	0.7%	0.2%	0.2%
forearms	0.0%	0.3%	0.1%	98.8%	0.1%	0.7%	0.4%	0.0%	0.1%	0.3%
lats	0.1%	0.9%	0.4%	0.3%	97.9%	1.5%	0.7%	0.1%	0.2%	0.5%
middleback	0.0%	0.2%	0.1%	0.2%	0.6%	94.9%	1.6%	0.1%	0.4%	0.2%
reardelts	0.0%	1.0%	0.5%	0.3%	0.4%	1.2%	95.2%	0.1%	0.8%	0.5%
shoulders	1.3%	0.1%	1.5%	0.0%	0.0%	0.1%	0.1%	97.4%	0.2%	0.1%
traps	0.2%	0.3%	0.2%	0.1%	0.1%	0.5%	0.6%	0.1%	96.8%	0.3%
triceps	0.1%	0.7%	0.3%	0.0%	0.3%	0.1%	0.8%	0.4%	0.7%	97.7%

Figure 3.21: Confusion Matrix of Exercise Classification (with $\boldsymbol{Study1\_univ}$ data)

3.6.3 Identifying the Exercise Performed

University Gym: We first evaluate the accuracy of classifying the 10 exercises (performed on the multi-purpose cable pulley machine) from 588 sets of data collected from 30 subjects in $\boldsymbol{Study1\_univ}$. Using the approach explained earlier in Section 3.5.4, we performed a 10-fold cross validation with Random Forest classifier and obtained a performance accuracy of **96.93%**, with a precision of 0.962 and recall of 0.969, in classifying the exercises. This is a mixed person model as it includes training data from all the users for all the exercises. From the confusion matrix (Figure 3.21), we found that the classification errors occurred primarily during *middleback*, *rear-delts* and *biceps* exercises, due to the higher *within-exercise* variability across users.

As expected, we obtain a higher accuracy of 98.75% in classifying the exercises when evaluated with the smartwatch data. From the smartwatch data, the misclassifications were mainly between (i) *traps* & *biceps* exercise and (ii) *shoulders* & *abs* exercise due to their similar range of arm movements. Table 3.11 further shows that the exercise classification accuracy is roughly comparable, for the wearable vs. weight stack sensor.

Using *InfoGainAttributeEval* in Weka, we further evaluated the features with the highest information gain. We utilized this attribute evaluator to study the worth of each feature by measuring the information gain (or in other words, how each fea-

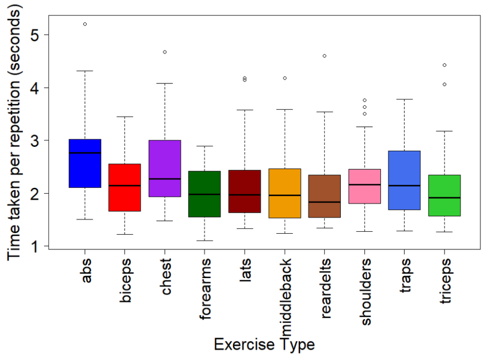


Figure 3.22: Average time taken to complete repetition per exercise across all subjects

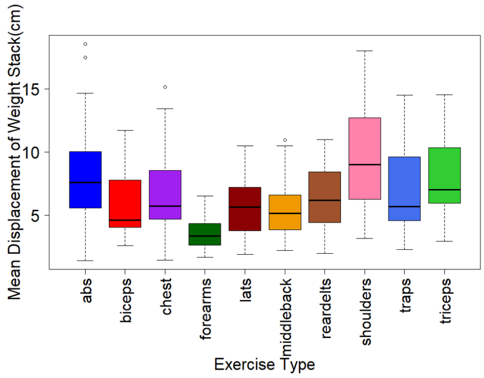


Figure 3.23: Average height to which weight-stack was lifted per exercise across all subjects

ture contributes in decreasing the overall entropy) with respect to the output class. A good attribute is the one with most information gain (i.e., reduces most of the entropy). We found that the *repetition-height* and *repetition-time* (both of which are derived from accelerometer data) were the most distinguishing features in exercise classification. To illustrate this, Figure 3.22 plots the distribution of the average time *per repetition* of each exercise across all 30 subjects. For most users, *abs* exercise took the longest time to complete ($\geq$2.65 secs), followed by chest exercise (approx. 2.3 secs). The triceps exercise took the least amount of time per repetition ($\leq$2 secs). Similarly, Figure 3.23 plots the boxplot of the variation of the height to which the weight stack was lifted for each of these 10 exercises. As mentioned earlier, the highest range of motion of the weight stack is for *shoulders* exercise and the least is for *forearms* exercise.

Community Gym: To further evaluate the exercise classification accuracy, we analyzed the $\boldsymbol{Study2\_comm}$ data, where users performed exercises using exercise-specific weight machines. We applied a 10-fold CV approach, where the data consisted of exercises performed across *all* the 6 machines. *W8-Scope* achieved an accuracy of 97.79% (precision=0.978, recall=0.982) in classifying the 6 exercises performed by 15 subjects. With a *leave-one-exercise-set-out* cross validation approach, the accuracy drops slightly to 94.4%.

Table 3.11: Performance of identifying the exercise performed – weight stack sensor data vs. smartwatch data using 10-fold cross validation

Exercise Classification	Accuracy	Precision	Recall
$\boldsymbol{Study1\_univ}$ Weight Stack Sensor Data	96.93%	0.962	0.969
$\boldsymbol{Study2\_comm}$ Weight Stack Sensor Data	97.79%	0.978	0.982
$\boldsymbol{Study1\_univ}$ Smartwatch Data	98.75%	0.987	0.986

3.6.4 Detecting Exercise Mistakes

For evaluating the performance of this component, we utilized the data collected for four variations of incorrect executions of two exercises (triceps and lats) from 30 subjects in $\boldsymbol{Study1\_univ}$. We also included the data from one set of *correct* execution for each exercise. The four incorrect variations included 3 explicit commonplace errors: {pulling too fast, releasing too fast, lifting only half-way} and one implicit error: "lifting too heavy a weight" (which is known to result in improper exercise dynamics). This last set was curated from subjective feedback provided by each participant, whenever they indicated that the *amount of weight* was 'too heavy' for them.

As explained earlier in Section 3.5.4, we extracted features on signals (*acceleration, velocity, displacement and magnetic*) corresponding to individual repetitions and labelled them with one of the four labels – {*Correct, Pull Fast, Release Fast, Lift Half Way*}, based on our ground truth. We first performed a multi-class classification to understand if we could distinguish between these three mistakes that are made while performing the cable pulley exercises. We obtained an overall performance accuracy of **96.75%** in classifying the mistakes. On performing a leave-one-subject-out cross validation (LOOCV), we observed a sharp drop in accuracy to 79.2% (precision=0.78; recall=0.82). The performance drop in LOOCV is explained by the fact that *mistakes are often person-specific*, with mistakes for one person appearing very similar to the correct execution by another user–e.g., the weight stack motion dynamics for a tall user *lifting half way* are very similar to a short user performing *correct lifting.* Given this observation, one possible approach

Table 3.12: Performance of identifying the mistakes made – weight stack sensor data vs. smartwatch data using cross validation

Mistakes Classification	Accuracy	Precision	Recall
10-fold CV using Weight Stack Sensor Data	96.75%	0.968	0.967
LOOCV using Weight Stack Sensor Data	79.2%	0.78	0.82
10-fold CV using Smartwatch Data	96.46%	0.965	0.965

to improve the mistakes classification performance could be to include some *user-specific* features (e.g., height of the user or body built). However, we did not have such information about the participants in our subject pool and thus, this is not explored further in this dissertation. Similar studies performed using the smartwatch data result in a mistake classification accuracy of 96.46% (Table 3.12 shows the detailed comparisons), indicating that the weight stack sensor is equally effective in capturing such typical mistakes.

3.6.4.1 Additional Insights into 'Typical Mistakes':

Because our long-term goal is to provide individuals actionable feedback to correct mistakes, we also performed manual annotation of the exercise videos (which provide "ground truth") to understand a few additional characteristics of such mistakes. Table 3.13 details the various fine-grained insights that we gained from this analysis.

1. *Does lifting 'too heavy a weight' result in disproportionately higher mistakes (e.g., 'releasing too fast', 'lifting only halfway', 'making postural mistakes')?*

   For this purpose, we manually annotated the 60 ground truth videos recorded, across 30 subjects, for the *triceps* and *lats* exercises performed with "heavy weight". The annotation was performed separately for the two individual transitions (upward and downward) of each repetition. We observed that, out of the 584 repetitions from 60 sets of lifting heavy weight, the subjects committed some kind of mistake (details listed in Table 3.13) during 93 repetitions across 21 sets (35% of heavy sets). Also, compared to exercise sets performed with lighter/comfortable weight, on an average the time taken to complete one repetition for *triceps* and *lats* exercises has also increased by 0.65 seconds. We used the previously trained *mistake-classifier* model and provided the data from the sets which had manually labelled 'mistake labels' as a test set. We obtained an overall accuracy of 81% (precision=0.84; recall=0.80) in classifying the two mistakes ('lift half way', 'release too fast'). In contrast, applying the same classifier to the 120 sets (of the same 30 users) which involved lighter weights resulted in the identification of mistakes in 57 repetitions across 14 sets (11.6% of non-heavy sets). This strongly suggests that mistakes in exercise motion dynamics are significantly more likely when gym-goers attempt to exercise with heavier weights.

2. *Are mistakes isolated (singletons) in a set, or do they consistently manifest across an entire set?* To answer this question, we randomly selected 10 subjects and manually annotated 197 videos of their 10 exercises performed naturally with two different weights ($\mathbf{3.75}\boldsymbol{kg}$, $\mathbf{6.25}\boldsymbol{kg}$). Out of the 197 exercise sets, 64 *reps* within 20 sets (10%) across 6 subjects had incorrect executions (i.e., had at least one rep with any of the 3 mistakes: 'pulling too fast', 'releasing too fast', 'lifting only halfway'). Moreover, *mistakes are often repeated*: 75% of the incorrect sets (15 out of 20) had 3-5 consecutive incorrect repetitions. Moreover, the *W8-Scope* classifier was able to correctly identify 83% of the mistakes performed in these manually-curated sets.

Key Takeaway: Our analyses suggests that *W8-Scope* can be used to reliably identify the majority of instances (repetitions) within an exercise set/session where a user makes commonplace "motion dynamics-related" errors. Such knowledge can then be used to tailor useful actionable feedback: e.g., observations of more frequent mistakes during *shoulders* exercise likely indicate weak shoulder muscles, and the gym-goer may be recommended additional shoulder exercises. However, our purely weight-stack based approach does not currently provide insights into other *postural*

Table 3.13: Insights into *Typical Mistakes* that people make – Observations from exercise videos

Key Observations	Supporting Evidence
People tend to make more mistakes while lifting heavy weights	35% of heavier weight lifted sets had mistake – lift half way (62 reps), release too fast (31 reps)
Postural mistakes such as "hunching the back", "leaning forward", "moving elbow during triceps exercise", "swinging body during lats exercise" are commonly made while lifting heavy weights	41% of heavy weight sets had mistakes with body postures – hunch (33 reps), lean forward (16 reps), move elbow (54 reps), swing body (67 reps),
People tend to mistakes constantly in an exercise set	75% (15 sets) of the incorrect sets had 3-5 consecutive reps that were incorrect
Most mistakes are made towards the end of an exercise set and in the second set of the same exercise	90% of incorrect sets have mistakes made from rep 6 and onwards
Lifting the weight half way through followed by releasing the weight too fast were the prominent mistakes	Out of 64 incorrect reps – lift half way (48 reps), release too fast (10 reps), pull too fast (6 reps)
Most number of mistakes were made while performing shoulders exercise followed by chest and abs exercises	Incorrect reps: Shoulders (53%), Chest (17%), Abs (12%)

mistakes that may be committed by novice users.

3.6.5 Identifying Users Performing Exercises

W8-Scope's final component helps to distinguish between the different users performing the same exercise. Table 3.14 summarizes our numerical results.

University Gym: Applying the 'User Classifier' across the 30 university gym users results in a classification accuracy (using 10-fold cross validation) of **98.97%**. Out of the 10 exercises, the classification errors are primarily confined to the *shoulders*, *forearms*, *middleback* and *triceps* exercises. On more careful inspection, we found that the users who were typically mis-classified had highly similar repetition-based features– i.e., having similar range of motion for the weight stack and taking the same amount of time to complete a repetition. By ranking the features based on its information gain, we found the most significant features to include: (a) *repetition time*, *displacement height* and *velocity* for the accelerometer sensor, and (b) *minimum*, *maximum* and *energy* of the 3-axes, for the magnetometer sensor. In contrast to *W8-Scope*, user identification using the wrist-worn smartwatch data provided a slightly higher accuracy of 99.31% (precision=recall=0.99). This is anticipated, as a wrist-worn smartwatch should be able to capture a greater range of arm motion, and thus acquire the exercise-specific movement differences across different users.

Community Gym: *W8-Scope's* 'User Classifier' achieves an accuracy of 98.74%

Table 3.14: Performance of user identification – weight stack vs. smartwatch

User Classification	Accuracy	Precision	Recall
$\boldsymbol{Study1\_univ}$ Weight Stack Sensor Data	98.97%	0.989	0.988
$\boldsymbol{Study2\_comm}$ Weight Stack Sensor Data	98.74%	0.985	0.987
$\boldsymbol{Study1\_univ}$ Smartwatch Data	99.31%	0.993	0.99

(precision=recall=0.98), when applied to the case of 15 users who performed 180 total sets of 6 different exercises. Note that the Community gym-goers were more diverse (in terms of various demographic factors and gym expertise). Our results thus demonstrate that *W8-Scope* can indeed be applied robustly to distinguish among users, across a wide variety of demographics.

3.6.6 Performance: W8-Scope vs Smartwatch Approach

Using the $\boldsymbol{Study1\_univ}$ data, we compared (and summarize in **Table 3.15**) the performance of each component of *W8-Scope* with that of an alternative smartwatch-based approach. Key results include: (a) A weight-stack mounted sensor is able to identify the weight lifted more accurately than a hand-worn sensor (overall accuracy of 84.37%, precision=0.822 & recall=0.845); (b) The smartwatch achieves slightly higher accuracy (98.75%) for exercise classification. (b) As expected, because of its ability to track the 3D arm motion precisely, the smartwatch has a slightly better accuracy of 99.31% (precision=recall=0.99) in identifying the user. (c) For identifying the exercise performed or any mistakes made, the performance of *W8-Scope* and the smartwatch is roughly comparable.

The overall accuracy of inferencing is high and comparable to that achieved by a wearable sensor based approach, demonstrating that a *cheap non-intrusive* weight stack sensor could substitute for a wearable sensor in accurately monitoring individual-specific exercise characteristics. Our evaluation also confirms the validity of the proposed approach across varying demographics of users and across different types of weight machines.

Table 3.15: Summary of performance accuracy – *W8-Scope* vs Smartwatch

	***W8-Scope* Approach** ($Study1\_univ$)	***W8-Scope* Approach** ($Study2\_comm$)	**Smartwatch Approach** ($Study1\_univ$)
Weight Classification	97.50%	N/A	84.37%
Exercise Classification	96.93%	97.79%	98.75%
Mistakes Classification	96.75%	N/A	96.46%
User Classification	98.97%	98.74%	99.31%

3.7 Medium Time-Scale Robustness: Adapting W8-Scope Classifiers

Results in Section 3.6 demonstrate *W8-Scope's* accuracy in tracking weight, exercise type and performing user. However, achieving these levels of accuracy in practice may prove to be elusive as our presented results were based on the use of training and test data from coterminous (or closely spaced in time) sessions. It is natural to ask whether *W8-Scope's* supervised learning approach will continue to provide high performance accuracy over medium-timescales (e.g., across weeks or months), especially as an individual's exercise pattern may be expected to evolve over such time periods. This *may* especially be a concern for exercise and user classification (which depend on the exercise-driven motion dynamics of the weight-stack), as opposed to weight determination (whose features are not really user-dependent).

To validate the robustness of our approach across exercise activities that are spaced weeks apart, we first use the data from first two sessions of $\boldsymbol{Study3\_long}$ (i.e., 10 users performing 5 exercises, across 2 different weeks) as the test set, applying our previously trained models with $\boldsymbol{Study1\_univ}$ data (i.e., from 30 users performing 10 exercises). (*Note:* As illustrated in Figure 3.24, $\boldsymbol{Study1\_univ}$ and $\boldsymbol{Study3\_long}$ are separated by a gap of over 3 months, with each of the 4 sessions in $\boldsymbol{Study3\_long}$ occurring in 4 consecutive weeks.) For these two sessions, we obtained an accuracy of 90.5% for weight classification, 78.3% for exercise classification and 75.2% for user classification, when the classifier outputs are ascertained *per-set* (using the "dominant-label" output across all the repetitions of an exercise set). This drop in classification accuracy especially for exercise (previously 96.9%) and user classification (previously 98.9%) suggest that a single-shot training of *W8-Scope* classifiers may indeed be inadequate in accommodating the evolutionary (medium timescale) changes in an individual's exercise patterns. To confirm that this accuracy loss is predominantly due to medium time-scale changes in individual exercise behavior, we trained new classifiers using the first two sessions of $\boldsymbol{Study3\_long}$ data and tested using the last two sessions. We obtained an accuracy of 93.1%, 89.8% and 90.4% for weight, exercise and user classification respectively, which are comparable to our results (on single sessions) in Section 3.6.

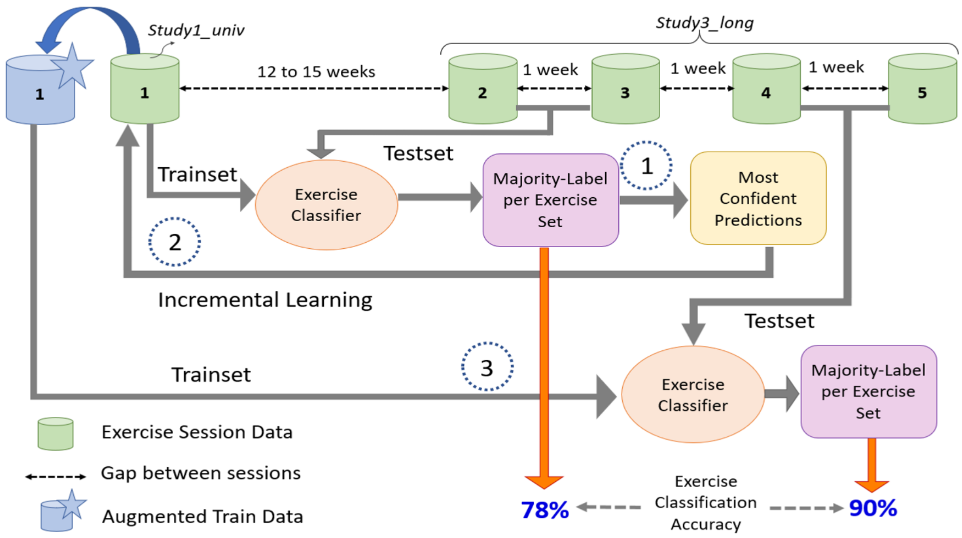


Figure 3.24: Incremental Learning with Longitudinal Exercise Data

3.7.1 Incremental Learning

To better incorporate such temporal evolution in the motion dynamics of individual exercises, we propose an enhanced *Incremental Learning-based W8-Scope* framework. Under this approach (Figure 3.24 illustrates the specifics, using "exercise classification" as an example, for our dataset), the labeled training data for the initially-trained *W8-Scope* classifier is continually augmented with those unlabeled

Table 3.16: Comparison of Medium Time-scale *W8-Scope's* performance (with and without incremental learning).

	Weight	Exercise	User
Without Incremental Learning	90.5%	78.3%	75.2%
With Incremental Learning	**95.1%**	**90.2%**	**87.4%**

exercise samples on which the classifier has *high confidence*. Very specifically, our *W8-Scope* instance starts off with the initial labeled training set (the $\boldsymbol{Study1\_univ}$ data). As an individual visits the gym, *W8-Scope* classifies the observed exercise activities, and then chooses the subset of such activity instances whose classification probability exceeds a given threshold $\boldsymbol{t}$. These "highly confident" samples are then used to augment the training set, and the classifier is retrained (on a per-weekly basis)–in Figure 3.24, step 2 (indicated within dotted circle) shows the process of augmenting the training data set with the 'highly confident' instances of recently-collected exercise data.

The performance of such incremental learning obviously depends on the right choice of the threshold $\boldsymbol{t}$. Intuitively, very low values of $\boldsymbol{t}$ will result in the addition of many noisy, likely mis-classified, samples in the training set. Conversely, very high values of $\boldsymbol{t}$ will lead to the incorporation of 'clean' samples, but might suffer from data paucity. Through empirical evaluation, we found that $\boldsymbol{t = 0.6}$ provides an appropriate choice between these two extremes.

3.7.2 Performance Results with Incremental Learning Strategy

We present below the changes in the performance of *W8-Scope*, after adopting this incremental learning strategy–i.e., the classification accuracy of activities performed during weeks 3 & 4 of $\boldsymbol{Study3\_long}$, based on a classifier augmented using 'highly confident' activity samples from weeks 1 & 2.

Weight Classification: The accuracy of weight classification was **95.1%**, with a precision and recall of 0.942 and 0.958 respectively. We observed that the classifier performance was poorer for certain heavier weights (e.g., $\boldsymbol{36.25kg}$, $\boldsymbol{43.75kg}$).

This is due to both the inability of a *single* magnetic sensor to perform fine-grained differentiation of heavier weights, as well as the lack of sufficient training data for heavier weights (most users exercise with lower weights).

Exercise Classification: We achieved an average set-level accuracy of **90.2%**, with a precision of 0.881 and recall of 0.923, in classifying the 5 exercises in the test set. When we analyzed the confusion matrix, we found that *biceps* exercise and *middleback* exercise were the ones typically mis-classifed, as they exhibited the greatest variability in the way these exercises were performed (e.g., high variance in repetition time, repetition height etc.) across various sessions and individuals, with individuals often also performing them incorrectly– e.g., not keeping the elbow fixed during *biceps curls* exercises.

User Classification: We achieved an accuracy of **87.4%** (with a precision=0.845 and recall=0.893) for discriminating among the 10 users (from a training subject pool of 30 total users) participating in $\boldsymbol{Study3\_long}$. The somewhat lower values of user classification accuracy were often due to *significant* changes in an individual's exercise style observed from the video feeds–e.g., when performing the *middleback exercise*, a subject initially used a bench to sit and perform the exercise, while in latter sessions, the user performed the same exercise while sitting on the floor and thereby altering the weight stack's overall range of motion. Based on our interactions with gym instructors, we gathered that such change in exercise behavior and postures are not commonplace for most individuals–i.e., this particular individual's behavior was likely an isolated case.

Table 3.16 shows the comparative performance of *W8-Scope* without and with incremental learning strategies. Overall, there was an increase of approx. 12% in the accuracy of classifying exercises and users after reinforcing the existing training set with highly confident samples from newly collected exercise data. These results suggest that as long as individual users visit the gym reasonably frequently (e.g., once every 1-2 weeks), *W8-Scope* can evolve its classifier models to capture the evolutionary changes in the individual's exercise motion dynamics.

3.8 Discussion Points

In this Chapter, I described the design and evaluation of *W8-Scope*, a system which can obtain quantified insights on various exercise-related attributes. We introduce a novel sensing mode (a combination of magnetometer & accelerometer) and sensor location (on top of a weight stack plate) for monitoring weight training exercises. Through extensive user studies conducted with 50 subjects in two real gyms, we consistently obtained an accuracy of 95%+ across all attributes, including the weight used, exercise performed, mistakes made and exercising user. We also show the need to adapt the classification model to accommodate real-world, longitudinal changes in user exercising behaviors, and show that an incremental learning-based approach provides sufficient robustness to our classifiers. Here, I outline additional preliminary investigations, extensions and open opportunities for this line of work.

Additional Sensors on Stack: In certain extreme cases, additional sensors on the weight stack may offer finer-grained discrimination. For example, we conducted an experiment with two sensors (one at the top and another at the bottom center of the weight stack), where an expert gym staff member performed *V-bar pull down* and *Seated cable rows* exercises with heavier weights on the cable pulley equipment. We collected data for 19 sets of 8 repetitions each for both the exercises, with the weights varied from 3.75kg to 48.75kg (which is the maximum weight). When we analyzed the data, we found that the magnetic sensor attached on top slab shows clear trend for individual repetitions until a weight equal to 38.75kg, whereas the *z-axes* of the magnetic sensor attached to the bottom showed discernible variation when a weight of 18.75kg or higher was lifted. Consequently, we observed that, across the *entire range* of weight slabs, the use of both top and bottom sensors results in a weight classification accuracy of **98.96%**, compared to 92.81% and 87.12% when one considers only the top or bottom sensor, respectively.

Extension to Additional Gym Equipment: Prior work [32] has shown the potential of using RFID tags attached to dumbbells to track the type and quality of dumbbell exercises. To study the possible application of the *W8-Scope* approach to other gym equipment, we conducted a small study with 4 users (2 sets, 10 reps) performing 6 different exercises (biceps curls, triceps extension, frontal raise, lateral raise, squats and lunges) using a sensor-attached dumbbell. By utilizing only the accelerometer sensor data, we obtained an exercise classification accuracy of 85%; however, user identification using this data proved more challenging. While these initial results look promising, we believe that additional strategies such as fusing data from wearable sensors or 3D tracking of the trajectory of the dumbbell would be required to track different exercise types, capture varying exercising styles and to scale to more number of users. This problem is further explored and discussed in Chapter 4.

Alternative Methods and Extensions: There are other recent works which tackles the similar problem of monitoring gym exercises of individuals. Bian et al. [19] introduced a wearable, body capacitance-based sensor for recognizing and counting seven different gym exercises. Unlike other wearable-based systems (which tracks only upper-limb exercises or uses multiple body-worn sensors), this system can track full body exercises just by using a single sensor attached to a body part which is not directly involved in the activity's movement. Guo et al. [41] uses WiFi CSI information to analyze workouts within a home/work environment. However, these WiFi-based systems may not work in a multi-user gym environment and in non line-of-sight scenarios. The GymCam [56] system leverages a single camera to track multiple people exercising simultaneously and recognize their exercise type and repetitions. However, this system does not identify the user and cannot track certain aspects of exercising such as the weight lifted or mistakes made.

Identifying incorrect body forms/postures: Weight training requires the user to adhere to specific exercise techniques as well as body forms/postures. Although our proposed approach can track incorrect exercise executions, it is not possible to infer the postural mistakes using only the weight-stack based sensor. To overcome this, we could extend *W8-Scope* by combining it with video-based contour tracking of participants, using either normal RGB or privacy-preserving thermal cameras. Such sensor fusion would allow us to track incorrect body postures and provide corrective feedback to prevent serious injuries.

Simpler Alternatives for User Recognition: While in this Chapter we present an approach for accurately recognizing the user performing a specific exercise, it is worth noting that there could be other easier alternatives to achieve this goal. The latest exercise machines come equipped with 'scanners' or 'keycode entry' (which is low-cost), which could allow for easier user identification.

Impact of Alterations to Gym Equipment: In this work we show that individuals' exercise behavior may evolve over time and such changes could be captured by approaches such as incremental-learning. Another factor that may possibly confuse our classifiers would be due to certain artefacts on the gym equipment itself. For example, replacing the cables of the exercise machine with newer ones may make it much stiffer, and consequently, it may affect the way individuals perform the weight training exercises. Additional investigations are required to better understand the impact of such practical situations and explore ways to accordingly fine-tune our approaches.

Interleaved Usage of Equipment: From field observations, we noted that weight stack machines occasionally saw "interleaved usage"–e.g., two users would perform their sets alternately. Our decision to perform exercise classification and user identification on a per-set basis are driven by this observation. In particular, we do not perform any additional 'majority voting' across sets. Of course, different users might also alter the settings of the weight stack during their exercises–such additional features might help to further improve our ability to discriminate among distinct users.

Enabling Near Real-time Analytics: In *W8-Scope*, all the analytics are performed in an offline manner. To obtain similar insights in a real-time manner, additional fine-tuning and tackling of system level challenges need to be done. An important aspect is to minimize the latency of the system in capturing various insights and enabling real-time suggestions. For example, currently even though *W8-Scope* performs 'mistakes classification' at a per-repetition level, it may not be feasible to provide the corrective feedback at the appropriate time (i.e., right after the repetition in which the person committed a mistake). However, a plausible approach would be to first detect all the mistakes in exercise execution and then give suggestions in a retrospective manner (e.g., the system could report that during *rep #4* and *rep #5*, the user performed the exercise only half way through) at the end of the set. Such kind of feedback would still be useful as individuals could take in this feedback and improve their performance during subsequent exercise sets.

3.9 Experiences and Lessons Learned

In this work, my research involved significant field experiments with real users in two gym facilities. Given these are 'semi-public spaces', there were several challenges in conducting clean experiments. These span from (i) recruiting various subjects of different expert levels (in weight training), age-group, gender, (ii) getting them to perform all the procedures and follow all the protocols as part of the user study, (iii) difficulty in using the gym equipment continuously for experiment purposes as it is also used in tandem by other gym users. These issues were more evident in the Community gym. In addition to these experiences, I also learned several lessons as part of conducting these user studies at the gyms. I briefly describe below some such lessons learned from this work.

- *User classification is hard with just IoT data*: Using only data obtained from non-personal IoT sensors (attached to exercise machine), it is difficult to distinguish between the different users performing various exercises on the same machine. Exercise behavior of individuals varied across sessions and often times the machine learning model was getting confused between users. Moreover, cross-validation approach ignores the behavioral evolution and does not help in accurate longitudinal tracking.

- *Practical difficulties in mounting sensors on machines*: While mounting sensors on the top of the weight stack was fairly easy and problem-free, securely mounting sensors on the bottom of the weight stack was more difficult. During the small scale studies conducted with sensors mounted on both the top and bottom of the weight stack, there were multiple instances where the sensor was falling off from the bottom of the stack. Also, at the community gym, mounting sensors even at the top weight slab was quite difficult as (i) the machine had a plastic outer shielding and inserting the sensor on the weight stack through the small opening was cumbersome and, (ii) the top surface of the weight stack was slightly curved and affected the firm placement of the sensor.

- *Difficulty in getting unusual data (e.g., heavy weights)*: Getting adequate data for exercises performed with heavier weights (e.g., 40kg or more) was difficult. Only the experienced gym staff or others who are experts in weight training could participate in studies involving such heavier weights. As such, most of our training data constitutes only those lower weights (with which majority of the participants could exercise with) and is not sufficient for accurate classification of all range of weights.

3.10 Acknowledgments

I would like to specially thank Prof Rajesh Krishna Balan and Prof Youngki Lee for their insights and help with drafting the initial survey distributed to SMU gym-goers. I'm also extremely thankful to all SMU gym staff and Toa Payoh ActiveSG gym staff for their valuable support and cooperation during the data collection phase. I would like to specially mention and thank Roy Chua, Gary Chua, Francis Lim from SMU, Huiyi Yu, Serene Yang from SportSG, Aron from Toa Payoh gym and Tony Varghese from GovTech for their kind help to coordinate and conduct all the user studies and to distribute our survey across ActiveSG public gyms as well as to staff members of SportSG and GovTech. I would also like to thank all the people who participated in the gym user study.

Chapter 4

Enabling Effective User Engagement during Gym Exercises

Extending on the work described earlier in Chapter 3, in this chapter I demonstrate the capability of using a popular, widely-used class of "in ear" devices to unobtrusively monitor weight-based exercises performed *concurrently by multiple individuals* in a gym. The work in this chapter is also motivated by feedback from users and potential technology adopters (e.g., executives from Singapore's SportSG government agency) that they would prefer solutions that went beyond just weight machines and covered other weight equipment (e.g., dumbbells and barbells). Compared to other class of wearables such as wrist, head or torso-worn devices, ear-worn devices have the added advantage of (a) being unobtrusive, (b) being widely used by gym-goers and (c) the ability to enable real-time audio-based feedback. I present a system based on a hybrid approach that utilizes sensor data from 'earables', fused with sensor data from non-personalized IoT devices attached to gym equipment (similar to as utilized in *W8-Scope* approach) to capture fine-grained exercising behavior of individuals. As I will show, this hybrid approach is necessary because: (a) due to its unfavorable in-ear location, it is very difficult to solely use earable devices to accurately infer the exercise-related limb movements, and (b) it is equally difficult to identify the individuals using a specific gym equipment, solely using sensors on such non-personal objects. In this chapter, I primarily discuss the system design and evaluation and also some key open opportunities that we are actively pursuing.

4.1 Monitoring Weight-based Gym Exercises of Multiple Concurrent Users

While there has been a rapid increase in the market for fitness devices and apps, relatively few solutions offer quantified and personalized feedback on an individual's overall exercise-related activities [59]. Also as discussed earlier in Chapter 3, existing technologies for fine-grained, individualized exercise tracking typically utilize video-based sensing [127], WiFi CSI information [41], or on-body wearable devices [25, 78]–each of these solutions continue to face challenges in real-world adoption. For example, video-based sensing generates significant privacy concerns, WiFi solutions suffer from poor accuracy in the presence of multiple individuals (e.g., at a gym) and individuals are reluctant to adopt custom wearable devices, *unless the wearable device is already a part of an individual's lifestyle*.

Motivated by these observations, we investigate the possibility of tapping on ear-worn (*'earable'*) devices (such as in-ear earphones) as a possible means of *capturing*, and, subsequently, *transforming* a user's exercise related activities. Earables offer a compelling and attractive *mass-market* wearable platform ([114] reported a global sale of 368 million headphones and headsets in 2018). Moreover, they are also commonly used during gym activities (e.g. for listening to music while working out). They also offer the advantage of supporting *real-time*, personalized *audio-based* feedback (often preferred to alternative text-based feedback [76])–for example, to rectify incorrect exercising behavior or to motivate continuation of desirable activities.

Key Challenge: The big drawback of earables, of course, is their unfavorable on-body placement: it is indeed questionable whether ear-based inertial signals can provide *any* discriminative information about exercise motion, especially when such motion is primarily restricted to upper or lower limbs. Research on earable-based activity recognition has been confined to inferring (a) characteristics of eating or drinking [18], both of which obviously manifest in head motion, and at a stretch, (b) high-level locomotive activities [82], which also involve overall body displacement.

In this work, I introduce a novel, low-cost solution for earable-based, individual-specific *fine-grained* monitoring of gym exercises in real world scenarios, *where multiple individuals are exercising concurrently*. Our key insight is that earable-based sensing, in isolation, is too noisy and weak to directly offer accurate recognition of gym activities. To overcome this limitation, we propose a hybrid architecture (to be elaborated in Figure 4.1), consisting of:

- Individuals wearing wireless earphones embedded with sensors (e.g., inertial sensors, heart rate sensor) that capture their activity and physiological context.
- Individual gym equipment (e.g., dumbbells, weight machine) attached with cheap IoT sensors that capture the motion dynamics of each equipment.

Given this architecture, the problem then morphs to (a) first establishing an *association* between an individual's earable device and the corresponding gym equipment, and (b) then using this pair of (earable, equipment) sensor data to infer fine-grained aspects of the exercise being performed. For the exercises performed on the weight machines, we allude to the point that *W8-Scope*-based user identification is not enough in this case. The idea here is that there may be multiple machines and different people are *simultaneously* exercising on different weight machines and in order to distinguish between concurrent users, such user-equipment association is required. While not part of this thesis, our overall vision also involves the generation of personalized real-time audio-based feedback (acting as a *"virtual personalized exercise coach"*), to the exercising individual, based on such fine-grained insights.

Using real-world studies conducted with multiple users concurrently performing weight-based exercises in a gym, we demonstrate the efficacy of the proposed

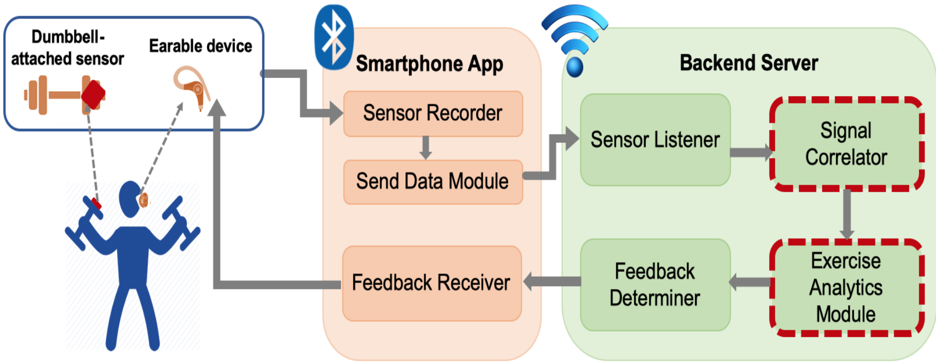


Figure 4.1: System Architecture

approach. Overall, our work also provides early evidence of the promise of earable devices as a platform for capturing fine-grained context of individuals exercising in a gym.

4.2 System Architecture

Prior works [62, 87] as well as our survey insights (described earlier in Section 3.1.3 of Chapter 3) reveal that gym-goers are interested in automatically tracking their exercise behavior and prefers to obtain personalized feedback on their performance. Past literature is, however, silent on the preferred timescales and frequency for such feedback–e.g., whether users would prefer to receive feedback *during* the ongoing repetitions in a set, at the end of individual sets or collectively at the end of an entire exercise session. As such, a future smart gym application should have the following capabilities: (i) distinguish between multiple people exercising simultaneously in the gym, (ii) unobtrusively monitor exercises performed by each individual and obtain deeper insights on various facets of exercising, (iii) provide personalized feedback to the individuals to improve the exercise effectiveness and prevent injuries.

For realization of such a smart gym application, we assume that individuals exercising in the gym are using *earables* and the exercise equipment/machine is attached with cheap IoT sensor devices. The earables are equipped with a microphone, inertial sensors (accelerometer, gyroscope), bio-sensors (heart rate, body temperature) and are paired to a smartphone. The IoT device attached to the exercise equipment (e.g., dumbbells, barbells, weight machines) have embedded accelerometer, gyroscope and magnetometer sensors. A custom built smartphone application has a *Sensor Recorder* process that records the sensor data from both the devices and a *Send Data* module that periodically transmits the sensor data to a backend server over the WiFi network. This App also has a *Feedback Receiver* that receives audio inputs/feedback from the server and relays it to the earables.

The backend server executes the required smart gym analytics components. In the backend, there is a *Sensor Listener* module for obtaining sensor data from both the earable and the equipment-sensor. Once the sensor data is obtained, the *Signal Correlator* module checks for the correlation between the earable sensor stream and equipment sensor stream to determine who is working out with which exercise equipment. The correlated sensor data pairs are then fed to the *Exercise Analytics* module, which identifies the type of the exercise performed and determines more fine-grained aspects such as the exercise intensity, correctness, heart rate variation for different exercises. Then, the *Feedback Determiner* module utilizes these analytics to determine the appropriate timing and the audio feedback to be sent to the earable device.

Figure 4.1 illustrates the architecture of the system with the sensor devices, server components and flow of the analytics pipeline. In this work, we mainly focus on the two components outlined in red-dotted lines. *Note:* For a clear representation, the figure depicts only a single-user scenario. In a practical setting, there will be multiple people exercising and thus multiple streams of both dumbbell and earable sensor will be streamed simultaneously to the backend sever.

4.3 Dataset

We conduct real-world studies, at our *University* gym, in which the participants performed a variety of weight-based exercises. The studies were approved by our Institutional Review Board (IRB-19-088-A078(919)). For the study, we recruited 12 (8 males, 4 females) university students and staff. Each study session involved multiple individuals performing exercises concurrently.

Sensor Devices Used: For obtaining sensor data, we used the following devices: (i) eSense Earable device[1], which the subjects wore on their left ear, (ii) Cosinuss One[2] earphone, worn by subjects on their right ear and (iii) a multi-sensor device (DA14583 IoT Sensor[3]) to attach to the exercise equipment (e.g., dumbbells, exercise machines). For the eSense earable, we used only the left-side earbud which has the capability to stream inertial sensor (accelerometer and gyroscope) data as well as receive audio inputs. The Cosinuss One device has in-built sensors to record heart rate and body temperature. These devices are paired with a smartphone and we developed an android application that simultaneously connects to these devices over Bluetooth Low Energy (BLE) and records sensor data and ground truth labels such as exercise performed, set count and amount of weight lifted.

Targeted Exercises: For the study, we focused on collecting data for 9 different exercises (listed in Table 4.1). This involved six free-weights exercises performed with dumbbells (both upper and lower body exercises) and three exercises performed on weight-based machines (we utilize a multi-purpose cable pulley machine).

Overall Study Procedure: Prior to data collection, the gym equipment (dumbbell and weight machine) was instrumented with the DA14583 IoT Sensor device. The subjects who consented to participate in the study visited the gym and they were first briefed about the study procedures. The participants were given the eSense

[1]eSense– `http://www.esense.io/`
[2]Cosinuss One– `https://www.cosinuss.com/products/one/`
[3]DA14583 IoT Sensor – (`https://www.dialog-semiconductor.com/iotsensor`)

Table 4.1: List of exercises and targeted primary muscle groups

Exercise Name	Primary Muscle Group	Exercise Equipment
Biceps Curls	Biceps	Dumbbells
Triceps Extension	Triceps	Dumbbells
Lateral Raise	Shoulders	Dumbbells
Side Bend	External/Internal Obliques	Dumbbells
Goblet Squats	Quadriceps	Dumbbells
Lunges	Glutes, Hamstrings, Quadriceps	Dumbbells
Standing Cable Lifts	Abs	Cable Pulley Machine
Bent Over Side Lateral	Shoulders	Cable Pulley Machine
Upright Cable Row	Traps	Cable Pulley Machine

earable (to be worn on their left ear) and the Cosinuss One (to be worn on their right ear).

A study session involved multiple users (varying from 2 to 4) who performed each exercise set concurrently. In a session, the subjects performed 3 sets of 10 repetitions of each of the 9 exercises. Note: for the cable pulley machine exercises, data was collected only when two people were exercising concurrently. Out of the three sets of each exercise in a session, the subjects concurrently performed the "same" exercise for 2 sets and for the last set, they alternated between "different" exercises. When performing each exercise set, all the subjects (exercising simultaneously) started exercising at the same time. However, the exercise set ending times varied depending on each individual's exercise pace. Overall, we collected 680 sets (of 10 reps each) of exercise data. All exercises performed by participants were video recorded for obtaining the ground truth. On an average, an exercise session per subject lasted for about 48 minutes. For participating in the study, we provided each participant a monetary compensation of $10. Table 4.2 summarizes the details of the user study.

Additional Small-scale Study: In addition to the actual user study, we also conducted a small-scale study at the gym to collect data for additional variety of free-weights exercises as well as for heavier weights. The main motivation for this study is to understand the role of earables in distinguishing between exercises with simi-

Table 4.2: Summary of real-world *multi-user* concurrent exercise dataset collected from University gym

	Study at University gym
No. of participants	12 (8 males, 4 females)
Age Variation	21-40 years
Self-rated expertise	5 (Novice); 7 (Intermediate); 3 (Expert)
No. of exercises	6 dumbbell exercises (3 upper-body, 3 lower-body) and 3 weight-machine exercises
No. of concurrent users	Concurrent user count varied from 2 to 4 2 users only (374 sets) 3 users only (162 sets) 4 users only (144 sets)
No. of sets of same/different exercise performed concurrently	Same exercise (452) Different exercises (228)
Total no. of exercise sets	680 sets (10 repetitions each)
Average duration of exercise session across subjects	48 minutes

lar dumbbell kinetics (e.g., Squats and Deadlifts) as well as exercises with different body postures (e.g., lying down for Weighted Crunch). For this study we recruited two people (in different sessions) who were well-experienced in weight-based training. In this session, they performed 6 different exercises namely, (a) Seated Barbell Shoulder Press, (b) Inclined Chest Flyes, (c) Dumbbell Triceps Kickback, (d) Weighted Crunch, (e) Barbell Deadlifts and (f) Alternating Bicep Curls. Compared to the previous set of dumbbell exercises which all had a "standing" posture, these exercises either have a "seated" or "lying down" posture or uses *barbells* instead of dumbbells. Both subjects performed 3 sets of 10 repetitions of each exercise. Additionally, they also performed 2 sets of 8 reps each of both *Biceps Curls* and *Lateral Raise* exercises with heavier weights (both 10kg and 14kg). In this study, we collected a total of 44 sets of data.

4.4 Earable-based Inertial Sensing for Exercise Activity Recognition

We focus on answering the following **key research questions:**

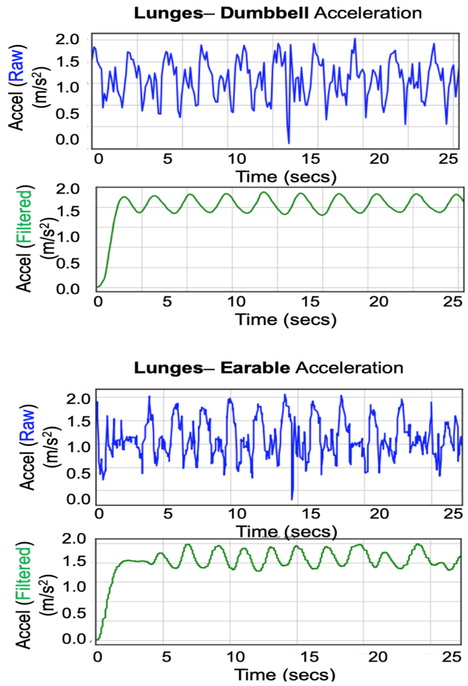


Figure 4.2: Raw (blue) and filtered (green) signals from Dumbbell (top) and Earable (bottom) for *Lunges* exercise

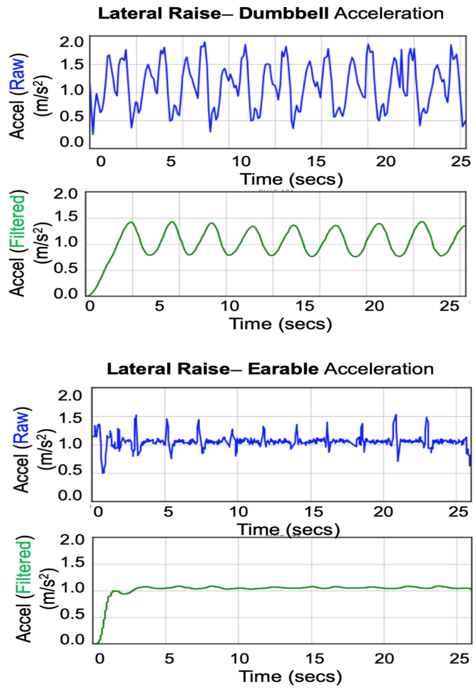


Figure 4.3: Raw (blue) and filtered (green) signals from Dumbbell (top) and Earable (bottom) for *Lateral Raise* exercise

- Does the accelerometer on the ear-worn sensor device show any discernible pattern for the common weight training exercises performed by individuals in a gym?
- Can we correlate the sensor data from the ear-worn device and the equipment-attached device to distinguish between individuals?
- Does the use of earable plus equipment data help improve the accuracy and/or robustness of exercise recognition?

We next describe our overall approach of analyzing sensor data and deriving various insights on the exercises performed concurrently by multiple individuals in the gym.

4.4.1 Sensor Data Analysis and Insights

We first inspect the accelerometer data recorded from the eSense left earbud and the equipment sensor. As expected, the equipment accelerometer showed clear and varying patterns for most of the exercises. For the earable, as any 'exercise-related' perturbations, if they exist, will be minor and may get swamped by various other macro-movements, we first pre-process and filter the sensor data. For this, we analyze the typical 'exercising frequency' of various exercises from the equipment sensor pattern. We observe that on an average the time taken to complete one repetition of a dumbbell/machine exercise is about $2 - 2.5$ seconds. As such, we use a fourth order Butterworth band pass filter with a lower cut off frequency of 0.4 Hz and a higher cut off frequency of 4 Hz to filter both streams of sensor data.

Figure 4.2 and Figure 4.3 shows sample plots of the magnitude of the raw and filtered sensor signals for *Lunges* and *Lateral Raise* exercises respectively. We find that exercises which involve larger body movements (e.g., *lunges*, *squats*, *abs* exercise on machine) exhibit clear patterns in the earable signal for each exercise repetition. However, for certain upper-arm exercises (such as *biceps curls*, *lateral raise*), variations are not clearly evident in the time-domain earable signal. This makes the problem both promising and challenging and requiring further analysis of both time and frequency domain of the signals.

As such, we propose to obtain *Continuous Wavelet Transform* (CWT) of the signals. We choose to use wavelet decomposition instead of other frequency domain techniques such as Fourier Transform or Power Spectral Density because of its ability to obtain both temporal and frequency resolution of the analyzed signal. Performing the signal decomposition at *both* temporal and frequency resolution is important for us to identify at "what" frequencies, variations occur in the signal and "when" it occurs, further for accurate inference of exercise-related motion dynamics. We use the *Morlet* wavelet and vary the scales from 1 to 100. Figure 4.4 plots the scalogram (which is the absolute value of the CWT coefficients of a signal,

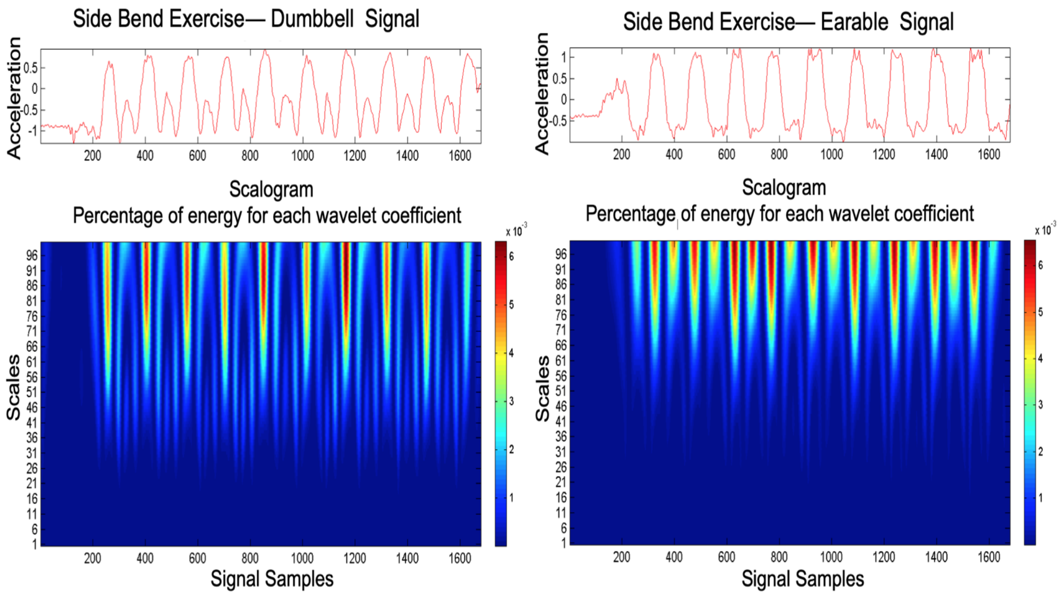


Figure 4.4: Continuous Wavelet Transform of Dumbbell (*left*) and Earable (*right*) Signal for *Side-Bend* Exercise

plotted as a function of time and scale) of one set of *Side Bend* exercise. From the figure, we can see that the individual exercise repetitions have their energy concentrated between scales 60 to 100. We observe similar trends for other exercises as well.

4.4.2 Identifying the Correct User-Equipment Associations

In our targeted gym scenario, multiple users would perform exercises simultaneously and the *smart gym* application should monitor exercise and provide personalized feedback to each individual. As such at the server side, we would receive multiple streams of both *earable* and *equipment* signals and therefore, our primary goal is to identify the correct pairs of $\{earable - equipment\}$ sensor streams to determine who is exercising with which equipment. Algorithm 1 outlines the steps taken to determine the association between the earable and equipment signals.

We first obtain the wavelet transform of the sensor signals. The CWT coefficients are computed at different scales for each of the filtered earable and equipment sensor streams. After performing CWT, we obtain a wavelet coefficient matrix

Algorithm 1 Association of Earable-Equipment Sensor Streams

Input: N_e *Earable Sensor streams* (E_1, E_2, E_3,...); N_d *Equipment Sensor streams* (D_1, D_2, D_3,...)
Output: $C_{Pairings}$: $\{E_1D_1, E_2D_2, E_3D_3, ...\}$
$fE, fD \leftarrow$ *NULL* {Initialize filtered signals list}
$scales \leftarrow range(1, 100)$ {Define scales for wavelet decomposition}
$W \leftarrow$ *NULL* {Initialize weights vector to hold pair-wise linear sum of feature values}
$C_{Pairings} \leftarrow$ *NULL*
for X, Y in N_e, N_d **do**
    $fE \leftarrow$ *bandpassfilter*(X, l$_c$, h_c) {Lower cut off, l_c=0.4Hz, Higher cut off, h_c=4Hz}
    $fD \leftarrow$ *bandpassfilter*(Y, l$_c$, h_c)
end for
for e in fE **do**
    for d in fD **do**
        $CWT_e \leftarrow$ *computeContinuousWaveletTransfrom(e, scales)*
        $CWT_d \leftarrow$ *computeContinuousWaveletTransform(d, scales)*
        $dCor_{ed} \leftarrow \frac{dCov(CWT_e, CWT_d)}{\sqrt{dVar(CWT_e)dVar(CWT_d)}}$ {Distance correlation between two CWTs}
        $P_E, T_E \leftarrow$ *segmentRepetitions(e)* {Segment repetitions and obtain peaks and troughs in signal}
        $P_D, T_D \leftarrow$ *segmentRepetitions(d)*
        for (p_{ei}, t_{ei}), (p_{dj}, t_{dj}) in (P_E, T_E), (P_D, T_D) **do**
            $pkGap_{ed}$ = *Distance*$\{p_{ei} - t_{ei}, p_{dj} - t_{dj}\}$
            $pkAlign_{ed}$ = *Distance*$\{p_{ei}, p_{di}\}$ + *Distance*$\{p_{ej}, p_{dj}\}$
        end for
        $W_{d,e} = \sum$(dCor$_{ed}$, μ_pkGap_{ed}, σ_pkGap_{ed}, $\mu\_pkAlign_{ed}$, $\sigma\_pkAlign_{ed}$) {Normalized linear sum of features}
    end for
end for
$C_{Pairings} = min \sum_d \sum_e W_{d,e} X_{d,e}$ {Hungarian algorithm to obtain the pairings}
=0

from both sensor streams of all exercising individuals. Next, we compute *distance correlation* between all the possible pairs of the coefficient matrices. The distance correlation is a measure of dependence between random vectors and is obtained by dividing the distance covariance of two matrices by the product of their distance standard deviations [120].

In addition to the wavelet features, we also compute temporal features such as: (i) *peak gap* and (ii) *peak alignment* from the earable and equipment sensor streams. For example, imagine a dumbbell time series that has a peak and trough (for 1 repetition) at times $\boldsymbol{td_1}$ and $\boldsymbol{td_2}$ respectively. Similarly, say $\boldsymbol{te_1}$ and $\boldsymbol{te_2}$ are the peak and trough of the earable signal. Our intuition behind the *peak gap* feature is that the time difference between peak and trough of the two streams (i.e., $\{\boldsymbol{td_1} - \boldsymbol{td_2}, \boldsymbol{te_1} - \boldsymbol{te_2}\}$) should match in case of the sensor streams that corresponds to an individuals' exercise set. In other words, a correct {equipment, earable} pair should have a low distance in such 'time domain distance' measures. Similarly, the *peak alignment* feature is based on our observation from the data that for majority of the exercises we considered, the peaks/troughs from individual repetitions occur at the same time on both the equipment and earable signals. Thus, the correct pair should have a lower peak alignment distance.

To compute these temporal features, we first perform repetition segmentation and counting on the signal (on the band-pass filtered sensor signals) and identify the crests and troughs of individual repetitions. The *peak gap* feature is defined as the distance between the crest and trough of the signal. We obtain a distance measure per repetition and compute the mean and standard deviation across all repetitions for both the equipment and earable signal. Then the difference of those peak gaps between all pairs of signals is computed as the final feature. The *peak alignment* feature determines if the peak time instants align between two sensor streams and is computed as the time difference between the peaks of each pair of equipment and earable signal. We compute the mean and standard deviation across repetitions for this feature as well.

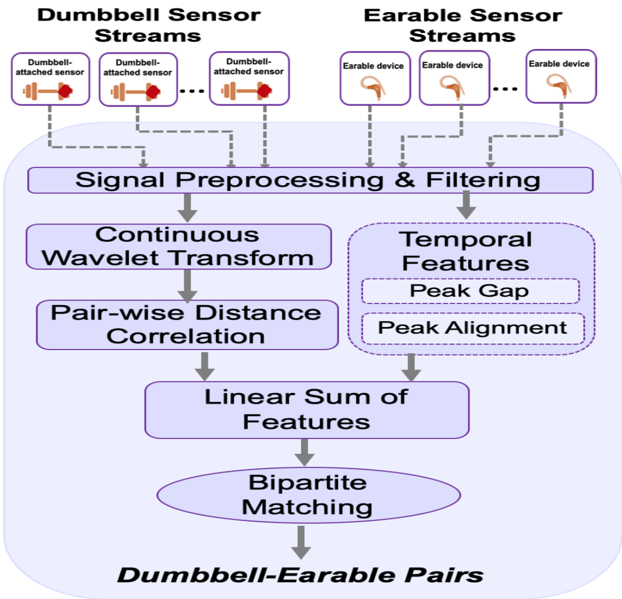


Figure 4.5: Steps involved in identifying the Correct $\{$*Equipment-Earable*$\}$ Pair

Once the wavelet and temporal features are extracted, we compute the normalized linear sum of these features. To identify the correct {earable-equipment} pair, we utilize the Hungarian algorithm [64] for bipartite graph matching. We assign the linear sum of features from different {earable-equipment} pairs as 'confidence scores' or 'weights' of the bipartite graph. i.e., Our matching logic is defined as a problem instance (described by a matrix $\boldsymbol{W}$), where each $\boldsymbol{W}[\boldsymbol{d}, \boldsymbol{e}]$ is the cost of matching vertex $\boldsymbol{d}$ of the first partite set (an "equipment") and vertex $\boldsymbol{e}$ of the second set (an "earable"). The goal is to find an assignment of 'equipment' to 'earable' of minimal cost. Formally, $\boldsymbol{X}[\boldsymbol{d}, \boldsymbol{e}]$=1 iff row $\boldsymbol{d}$ is assigned to column $\boldsymbol{e}$. Then the optimal assignment has cost:

$$\boldsymbol{C} = \boldsymbol{min} \sum_{\boldsymbol{d}} \sum_{\boldsymbol{e}} \boldsymbol{W}_{\boldsymbol{d},\boldsymbol{e}} \boldsymbol{X}_{\boldsymbol{d},\boldsymbol{e}} \tag{4.1}$$

such that each row is assignt to at most one column, and each column to at most one row. The advantage of this 'matching' technique is that it can be easily extended to scenarios with a larger number of concurrent users and also incorporate practical situations where all exercising individuals may not be wearing earable devices (for example, in such cases, we would obtain $\boldsymbol{M}$ earable and $\boldsymbol{N}$ dumbbell sensor

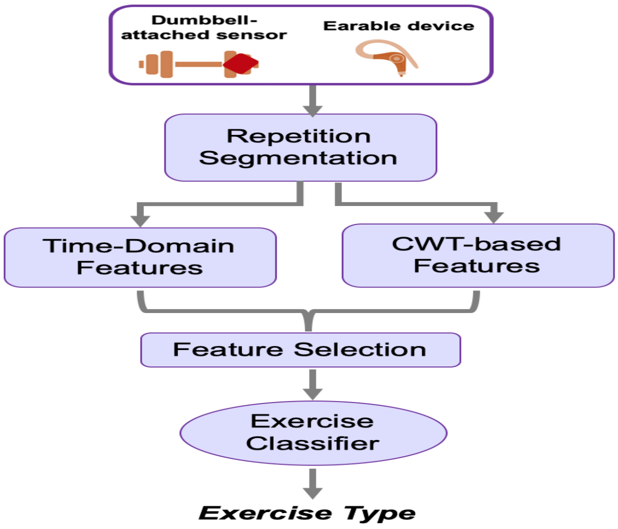


Figure 4.6: Steps involved in identifying the Exercise Type

streams ($M < N$)). The classic assignment problem can be generalized to such scenarios using *inexact bipartite matching* techniques (where the cost matrix is rectangular). Figure 4.5 depicts the various steps involved in determining the correct $\{$equipment-earable$\}$ pairs.

4.4.3 Identifying the Exercise Performed

Once the correct equipment-earable pairs are determined, the next step is to identify the exercise performed by each individual. We use a supervised Random Forest (RF) classifier trained on features extracted from both the equipment and earable signals (Figure 4.6 depicts the various steps involved). The features are extracted on signals corresponding to individual repetitions from both the earable and equipment sensors. The features computed include statistical time and frequency domain features (similar to as proposed in [135]) on the signal as well as the continuous wavelet transform of the signal. We then utilize correlation-based feature subset selection [44] to determine individual predictive ability of each feature and the degree of redundancy between them. We then train the RF classifier on a subset of features that are highly correlated with the output class while having low inter-correlation values. We also train models with only features extracted from either the equipment

Table 4.3: Summary of performance evaluations conducted.

Evaluation	Importance
Performance of Association	
Multiple concurrent users (Section 4.5.1)	To study the correct user-equipment association for different number of concurrent users (N={2,3,4}).
Same vs Different exercises (Section 4.5.1.1)	To study association performance when 'same' vs 'different' exercises are performed concurrently.
Association features (Section 4.5.1.2)	To study the discriminatory power of different features and its combination.
Different types of exercises (Section 4.5.1.3)	To understand the variation in association performance per exercise.
Amount of repetition data (Section 4.5.1.4)	To get a sense of the feasibility and robustness of early, real-time recognition and corrective feedback.
Inexact association (Section 4.5.1.5)	To capture real-world environments where there is a mix of people with and without earables.
Different start times (Section 4.5.1.6)	To understand association performance in real-life scenarios where users don't actually start exercising at the 'exact' same time.
Performance of Identifying Exercise Performed	
Equipment only vs Earable only (Section 4.5.2.1)	To study performance of exercise classification when only either of the equipment or earable sensor data is available.
Other exercise types (Section 4.5.2.2)	To study exercise classification performance for other kinds of free-weights exercises.

sensor or the earable to investigate the efficacy of using earable + equipment data for improved accuracy and/or robustness of exercise recognition. This is explained later in Section 4.5.2.1.

4.5 Real-World Evaluation

In this section, I present the performance evaluation of our approach, along with insights gained, based on real world exercise data (described earlier in Section 4.3) collected from our University gym. Table 4.3 outlines a high-level overview of the different performance evaluations conducted and the significance of those.

4.5.1 Performance of Identifying the Correct User-Equipment Pairs

We first evaluate how our proposed approach of associating the correct earable with the equipment (e.g., dumbbell) performs with naturalistic real world multi-user gym data. We study the association accuracy for varying number of *concurrently* exercis-

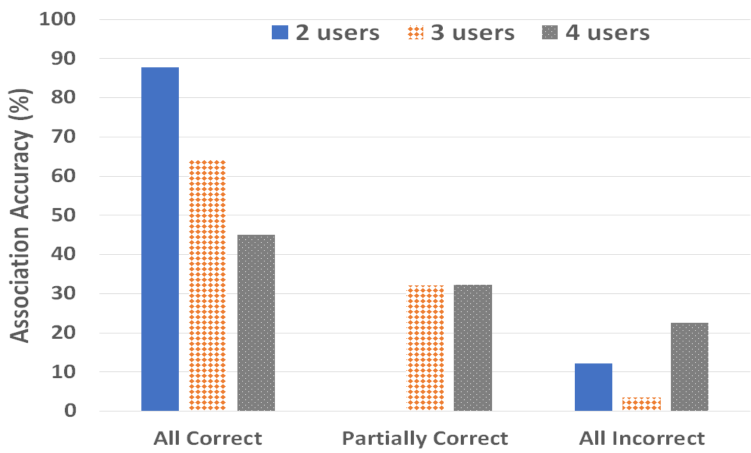


Figure 4.7: Overall accuracy of association when N={2,3,4} people perform exercise concurrently

ing individuals (i.e., with all N individuals starting each set simultaneously). Note again, for our study, we varied N from 2 to 4. Using the 680 sets (i.e., 374, 162 and 144 sets of N={2,3,4} concurrent users respectively) of data collected, we obtained an overall association accuracy of 88%, 65% and 45% in identifying *all* the correct pairs when N={2,3,4} respectively. We also looked at the performance of correctly identifying at least 1 pair (for 3 user cases) and at least 1 or 2 pairs (for 4 user cases). Our approach obtained an accuracy of 32% each in *partially* identifying the correct pairs. So, overall 12.2%, 3.5% and 22.5% cases of N={2,3,4} had all the pairs incorrectly identified. Figure 4.7 plots the performance accuracy of our approach. We performed further analysis to understand the impact of different factors such as the exercise-specific characteristics, importance of specific features.

4.5.1.1 Association Accuracy for Same vs Different Exercises Performed Concurrently

We next investigate if there is any notable difference in the association accuracy when people were performing *same* vs *different* exercises concurrently. For this purpose, we analyzed the association accuracy separately for the exercise sets belonging to these two categories. Figure 4.8 and Figure 4.9 plot the association accuracy for “same” exercise set and “different” exercise sets. We found that the associa-

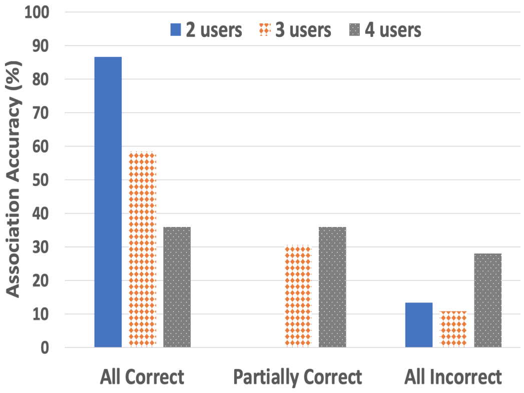


Figure 4.8: Association accuracy when N={2,3,4} people perform the *same* exercise concurrently

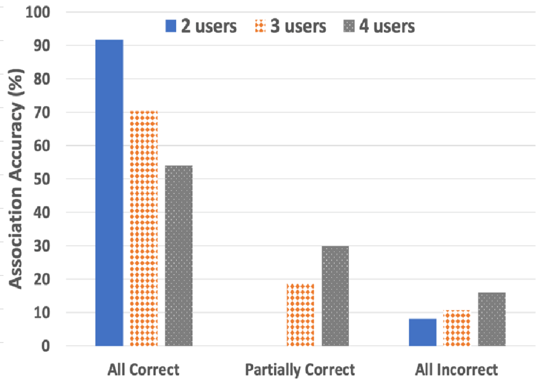


Figure 4.9: Association accuracy when N={2,3,4} people perform *different* exercise concurrently

tion accuracy was significantly better (especially for 3 and 4 user cases where there was an improvement of $>$10%) when people were performing *different* exercises together.

4.5.1.2 Discriminatory Power of Wavelet vs Temporal Features

We next study the discriminatory power of the two category of features: (a) wavelets and (b) temporal in determining the correct user-equipment pair. We obtain the association accuracy for cases when only wavelet features and only temporal features were used. We observe (see Figure 4.10) that *wavelet*-based features have slightly better discriminative power (10% higher accuracy) than the *temporal* features. This study also confirms that combining both set of features helps in significantly improving the performance accuracy.

For the continuous wavelet transform (CWT) features, we further investigated if any specific scales of the CWT has higher predictive ability. We computed the association accuracy across individual scales of the CWT and found that scales in the range of 70-100 obtained the highest performance (as shown in Figure 4.11). This shows that CWT at lower scales did not have any information value and performing wavelet decomposition only for higher scales (i.e., lower frequencies) was enough to obtain the same accuracy. We also observe that, on an average, the correct pair

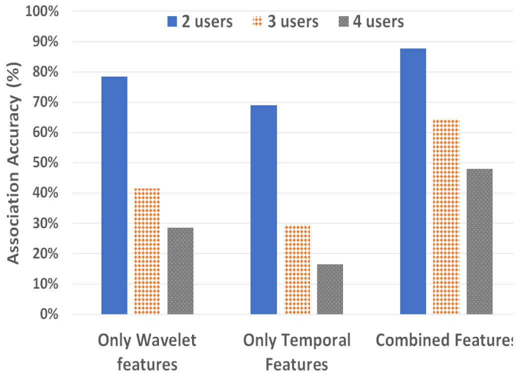


Figure 4.10: Association accuracy for different features

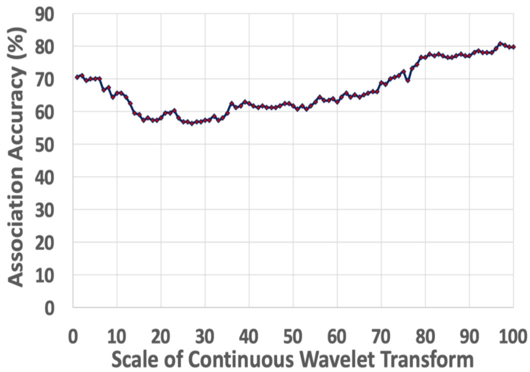


Figure 4.11: Association accuracy for different scales of CWT

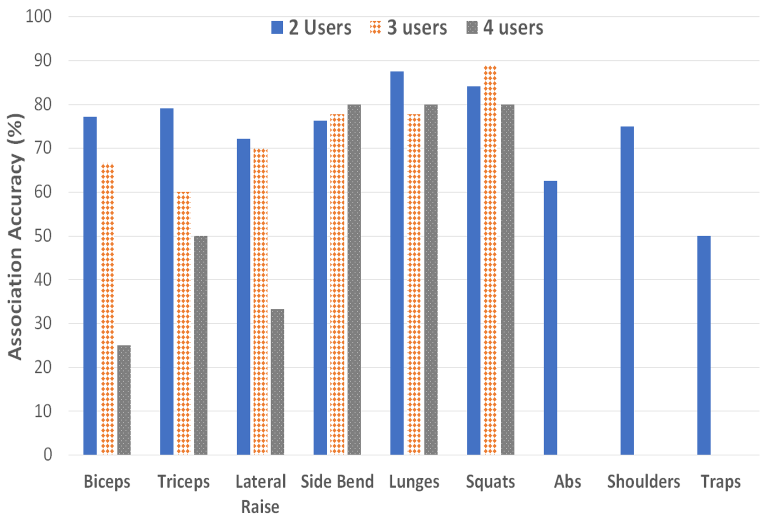
of signals (i.e., from same user's dumbbell and earable) have high correlation value over **0.71**.

4.5.1.3 Association Accuracy for Different Exercises

To understand the variation in performance accuracy for different type of exercises, we obtained the pairing accuracy for each of the 9 exercises independently and plot it in Figure 4.12. We observe that the 3 upper-body exercises (namely, biceps curls, triceps extension and lateral raise) had the lowest performance especially when the number of concurrent users were more than two. This is primarily because these exercises involve very limited head motion and the earables does not clearly capture exercise motion dynamics unlike the lower-body exercises like lunges or squats.

4.5.1.4 Impact of the Amount of Repetition Data on Association Performance

As our broader goals involve providing real-time feedback to the individuals while they are exercising, we next study the impact of amount of exercise repetition data required for accurate pairing. In Figure 4.13, we plot the association accuracy by varying the number of repetitions data used for pairing from 1 to 10. When two users are exercising concurrently, we found that the number of repetition instances did not have any significant impact on the overall performance achieved, showing that early, real-time pairing is possible. However, in cases of 3 and 4 concurrent exercisers,



Note: The association performance of *Abs*, *Shoulders* and *Traps* exercises (performed on weight machine) were evaluated only for two user scenarios as the weight machine could only be used by two people concurrently

Figure 4.12: Association accuracy across the 9 different exercises for N={2,3,4} people exercising concurrently

there was a notable drop in accuracy if only initial few (1 to 4) repetitions' data was used for our matching algorithm.

To further understand the cause for a drop in association accuracy when only data from initial few *reps* were used, we inspected the data for some cases where the 'associations' were incorrect. We observe that as repetitions progresses, there is divergence in time synchronization even when all concurrent users starts exercising at the exact same time. This causes the temporal features (e.g.,*peak alignment*) of the concurrent users to be very similar during the initial few repetitions, thus

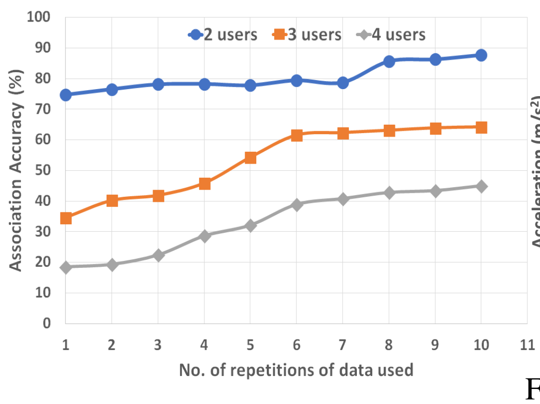


Figure 4.13: Amount of repetition data used vs Association accuracy

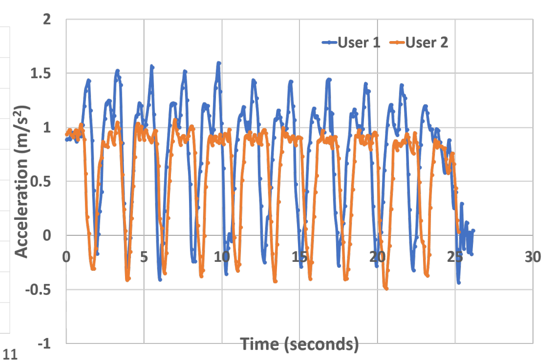


Figure 4.14: Acceleration patterns of two users concurrently performing *lateral raise* exercise.

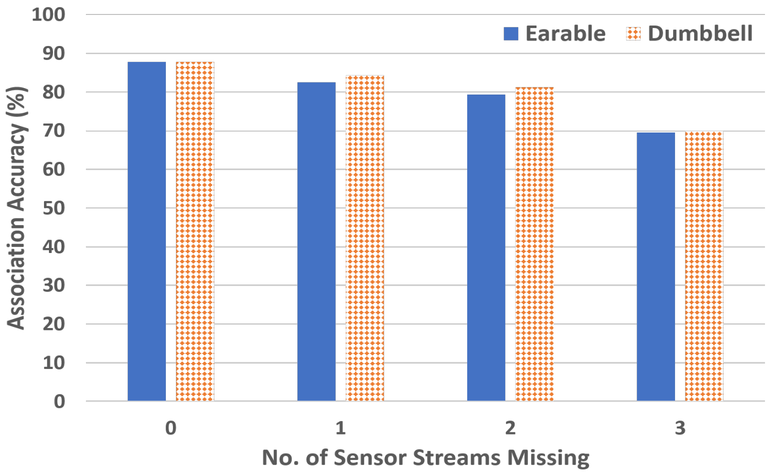


Figure 4.15: Association accuracy with Inexact Matching

confusing the association logic when only data from first few reps are used. Figure 4.14 shows a sample time series of the dumbbell data for two users concurrently performing *lateral raise* exercise where the *rep alignment* starts to diverge after few initial reps.

4.5.1.5 Association Accuracy with Inexact Matching

In practical situations, there may be cases where not all exercising individuals would be wearing an earable device or not all gym equipment is attached with a sensor device. We evaluate the robustness of our matching technique (with inexact bipartite matching) in such scenarios. To study this, we randomly removed 1 to 3 streams of earable *or* dumbbell signals at a time and obtained the association accuracy. Figure 4.15 shows the comparison of the association accuracy for the cases when either the earable or dumbbell was "missing" to that of "exact" matching (i.e., when no sensor streams are missing). We observe that association accuracy is $\approx$ 80% or more when only one or two sensor streams are missing. However, when three of either the earable or dumbbell sensor streams are missing, the accuracy drops to about 69%. Overall, there was a drop in accuracy of about 5-18.5% when one to three streams of 'earable' data was missing and a drop of about 3-18% when one to three streams of 'dumbbell' data was missing.

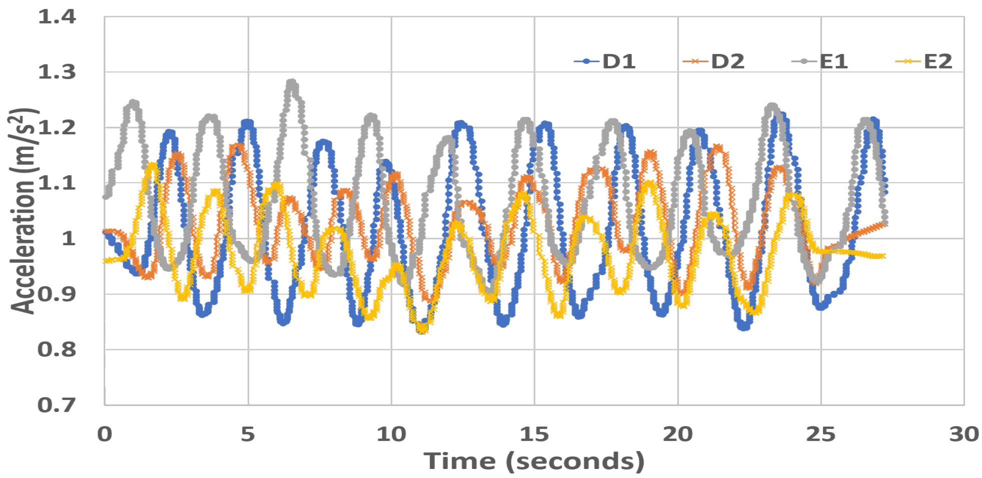


Figure 4.16: Dumbbell (D1, D2) and Earable (E1, E2) accelerometer sensor patterns of 2 users concurrently performing Squats exercise.

Table 4.4: Cost matrix for bipartite matching obtained based on the linear weighted sum of the features (obtained from the 4 sensor streams of two conncurent users) as weights .

	D1	**D2**
E1	0.54	0.84
E2	0.72	0.69

As an illustrative example, we show a case where our association approach failed with inexact matching (i.e., either of the dumbbell or earable sensor streams were missing). Figure 4.16 plots the dumbbell (blue and orange lines) and earable (grey and yellow lines) accelerometer time-series of *two* subjects concurrently performing *squats* exercise. In this case for both subjects, we observe that at the beginning of the exercise set, the peaks in the 'earable' signal occur slightly before the 'dumbbell' signal and then gradually aligns with each other towards the end. Table 4.4 shows the cost matrix obtained based on the linear weighted sum of wavelet-based distance correlation and temporal features. Using our 'perfect' association logic (i.e., when all four sensor streams are present) and with *Hungarian* assignment with (that minimizes the weights), $\boldsymbol{D1}$ is correctly matched with $\boldsymbol{E1}$ and $\boldsymbol{D2}$ is matched with $\boldsymbol{E2}$. However in the case of 'inexact' matching (when one of the signals is missing), the association logic gets confused and is unable to correctly identify the right pair. For example, in the simulation when $\boldsymbol{E2}$ (yellow line) was missing, instead of matching $\boldsymbol{E1}$ (grey line) with $\boldsymbol{D1}$ (blue line), it was incorrectly associated with $\boldsymbol{D2}$ (orange line). We found that $\{\boldsymbol{D2}, \boldsymbol{E1}\}$ pair had lower values of 'time domain distance' measures (i.e., the mean peak gap and mean peak alignment values) compared to the correct $\{\boldsymbol{D1}, \boldsymbol{E1}\}$ pair, leading to incorrect association.

4.5.1.6 Association Accuracy for Different Start Times

In our user study, the concurrent exercisers were asked to start exercising at the exact same time. However, this is an artificial and pessimistic scenario: in reality, even concurrently exercising users are likely to start their sets with slight time differences. To understand the performance of our approach in such situations, we simulate such a situation by perturbing the start time of each individual by small $\boldsymbol{\Delta}$ values. The value of $\boldsymbol{\Delta}$ varied from 0 to 2 seconds (as an individual exercise repetition takes about $\approx \mathbf{2}$ seconds) with steps of $\mathbf{0.5}$ seconds. Then for the experiment, for each $\boldsymbol{\Delta}$, we choose different values uniformly from [0, $\boldsymbol{\Delta}$] to shift the start times. We repeat this experiment 5 times with different random seeds.

Figure 4.17 shows the accuracy of identifying all the distinct pairs for different values of start time perturbation of each individual in a group. We found that with increased variation in the start time of different concurrently exercising individuals, the association accuracy improves slightly. While there was only an increase of $\approx$4% in the case $\boldsymbol{N}$=2 users, the improvement in association accuracy was higher for $\boldsymbol{N}$=$\{$3,4$\}$–i.e, up to 6% for $\boldsymbol{N}$=3 and up to 9% for $\boldsymbol{N}$=4 concurrent users. With a perturbation of 2 seconds in starting time, incorrect-pairing percentages for $\boldsymbol{N}$=$\{$2,3,4$\}$ cases were 7.6%, 28.6% and 45.7% respectively. With this experiment, more number of lateral raise exercise sets became correctly associated. However, both *biceps* and *triceps* exercise sets continued to have several incorrect pairings. Similar to our previous observation, the poor performance for these exercises are due to the inability of the earable device in capturing the accurate exercise motion pattern, which is because of the minimal head movements involved while perform-

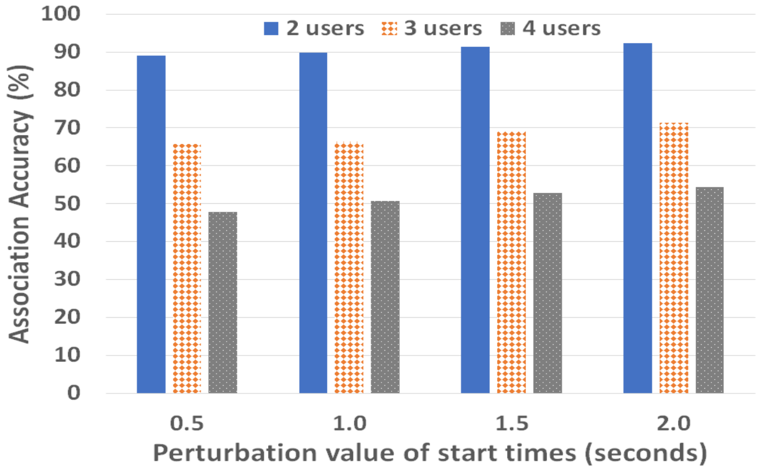


Figure 4.17: Association accuracy with perturbed start times of concurrent exercising individuals

ing these exercises.

4.5.2 Performance of Identifying Exercise Performed

We next evaluate the accuracy of classifying the 9 exercises (dumbbell and machine exercises). We divided the dataset into train (75%) and test (25%) sets. As explained earlier in Section 4.4.3, we trained a Random Forest classifier and performed 10-fold cross-validation on the training dataset. We obtained an accuracy of 92.3% with a precision and recall of 0.92 and 0.919 respectively in classifying the exercises. We then utilized the 10-fold cross-validated model and supplied the test set. On the test set, the model achieved an accuracy of 85.2% (precision=0.856, recall=0.849) in classifying the exercises. Figure 4.18 shows the confusion matrix of exercise classification. On inspecting the confusion matrix, we found that the classification errors occurred mainly for the following exercises: *triceps pushdown* & *lateral raise* (dumbbells) and *bent over side lateral* (shoulder exercise on machine), which has comparatively lesser head movements involved. 'Lunges' exercise achieved the highest performance with a precision of 0.944 and recall of 0.956.

Predicted Label	Actual Label								
	Biceps	Triceps	Lat-Raise	Side-Bend	Squats	Lunges	Shoulders	Abs-Lift	Traps
Biceps	**0.85**	0.15	0.03	0	0	0	0	0	0
Triceps	0.08	**0.69**	0.12	0	0	0	0.03	0	0
Lat-Raise	0.05	0.08	**0.74**	0	0	0	0.04	0.01	0.03
Side-Bend	0	0.02	0.07	**0.94**	0.04	0.02	0.02	0	0
Squats	0	0.01	0.02	0	**0.90**	0.03	0	0	0
Lunges	0	0	0	0	0.06	**0.95**	0	0	0
Shoulders	0	0.03	0	0	0	0	**0.79**	0.08	0.01
Abs-Lift	0.02	0.01	0.02	0	0	0.01	0.08	**0.89**	0.04
Traps	0	0.01	0	0	0	0	0.04	0.02	**0.92**

Figure 4.18: Confusion Matrix of Classifying the 9 Exercises

Table 4.5: Performance of identifying exercise performed for different modalities of sensor data used

	Both Sensors	Only Equipment Sensor	Only Earable Sensor
10-fold CV Accuracy (%)	**92.3%**	86.84%	60.46%
Test set Accuracy (%)	**85.2%**	78.63%	54.56%

4.5.2.1 Exercise Classification with Only one of the Sensor Modality

We next investigate the performance of exercise classification when only *either* of the dumbbell or earable sensor data is used to train the model. On the test set, we obtain an average accuracy of *78.63%* and *54.56%* by considering only dumbbell or only earable data, respectively. Although the average accuracy obtained with earable is quite low, we observe that earable-based sensing has higher predictive power in identifying certain exercises. For example, the precision of *lunges* exercise is $\approx \mathbf{0.78}$ with earables. Whereas the precision with just dumbbells is only $\mathbf{0.65}$ and with combined features from both sensors the overall precision for lunges significantly increases to $\mathbf{0.94}$. This shows that combining both dumbbell and earable data helps to increase the performance of classifying exercises. Table 4.5 presents the summary of results.

We also obtained the classification accuracy for the weight stack-based exercises using just (a) earable and (b) IoT sensors. Note that the model is trained with data from all 9 exercises (6 dumbbells, 3 weight machine exercises). The three weight stack-based exercises (targeting abs, shoulders, traps muscles) obtained an accuracy of 59%, 45% and 52% with *only earable data*. With just the IoT sensor (on the weight stack), the model achieves an accuracy of 86% for abs, 77% for shoulders and 90% for traps exercises. We observe that there is a notable drop in the accuracy, especially for the *abs* and *shoulders* exercise, compared to that we achieved with the *W8-Scope* approach (98%, 97% and 96% respectively for these three exercises as described earlier in Section 3.6.3 of chapter 3). This drop in accuracy is due to two main reasons: (i) in this approach we leverage only the accelerometer data, while *W8-Scope* uses a combination of both accelerometer and magnetometer sensor data for exercise classification and, (ii) we did not use the weight stack displacement-based features (e.g., the height to which or the speed with which the weight stack was lifted) in this case.

The purpose of leveraging magnetic sensor data in *W8-Scope* approach was to mainly determine the amount of weight lifted. This was not applicable in the case of dumbbells and moreover, it would have been more useful to have a magnetometer in the ear-worn devices. When the 'earable' hardware evolves with additional sensing capabilities, it will be worth to explore if there is any added advantage in using magnetic sensor data in more accurately distinguishing the different exercises. Besides, we could not directly compute displacement-based features for dumbbell exercises and we utilized a common model to distinguish between all exercises (both dumbbells and weight machine exercises). For obtaining the range of motion of the dumbbells, additional strategies such as 3D tracking of its trajectory would be required.

4.5.2.2 Classifying other Variety of Exercises

With the additional data collected from a set of different free-weights exercises (described in Section 4.3), we evaluate the performance of our exercise classification approach. To the dataset, we also included data corresponding to the 6 dumbbell exercises (from the user study) by randomly choosing equal samples to obtain a balanced data set of 12 different exercises. Using 10-fold cross validation, we obtain an accuracy of 89.2% with a precision of 0.90 and recall of 0.88 in classifying the 12 exercises. Among these exercises, *shoulder press* and *deadlifts* had the lowest performance of 80% and 77.2% respectively. For example, the 'Deadlifts' exercise was getting confused most with 'Squats' exercise as their exercise motion dynamics were very similar. Overall this evaluation shows that our approach is robust in tracking a variety of weight-based exercises. This study also confirms that we need both earable plus equipment data to distinguish robustly across a wide variety of exercises as using *only dumbbell* or *only earable* achieved an accuracy of just 75% and 52% respectively.

4.6 Discussion Points

While initial results are promising, there are several other aspects and open questions that we are actively pursuing to make this vision an eventual reality. We discuss below some of those key points as well as observational takeaways based on the real-world studies conducted.

Real-time audio feedback: Providing personalized feedback on the exercise progress and correctness could help improve exercise effectiveness as well as retain motivation to continue exercising. Prior studies [76] have reported that "auditory feedback" is ranked top among feedback features based on a review of physical activity apps. Based on real-time sensing and analysis of the multi-modal sensor data, we intend to provide incremental feedback in the form of short audio instructions or "beep" sounds, based on the performance and progress of users. The system could also provide positive, motivating feedback after completing each exercise set and at the end of the gym workout for the day. Motivated by prior work [82], we shall also investigate if we can use music to regulate the 'exercise tempo' of users.

Integrating physiological sensor data: Besides inertial sensor data, additional physiological signals (e.g., heart rate or breathing rate) from earables may enable more sophisticated monitoring or intervention strategies. For example, the physiological data could reveal the user-perceived intensity of the current exercises and enable the delivery of appropriate corrective feedback–e.g., alerting the user to slow down if the heart rate exceeds the maximum active heart rate. More interestingly, such physiological signals may provide additional temporal markers for better matching of {earable-equipment} pairs, especially for exercises with imperceptible head motion–e.g., if the inhalation/exhalation times match with the exercise repetition dynamics.

Improving performance with additional sensors: Future earable devices may also come equipped with a magnetic sensor. We thus pose the question: "Can the in-ear magnetic sensor represent any signal variations that may occur when the exercising equipment (e.g., dumbbell) is brought towards the ear while performing certain exercises (e.g., biceps curls)?". If it does, then fusing the magnetic sensor data with accelerometer data would help in improving association accuracy as well as in more accurately identifying certain exercise types. Similarly, an in-ear vibration sensor may help to capture very weak and minute head movements that are specific to certain exercises.

Extending to other exercise types and scenarios: In this work, we focus only on weight training exercises (both free-weights and machine weights). However, we believe that the ear-worn sensing platform can be used to monitor other types of gym exercise (e.g., cardio, body-weight exercises) and other outdoor exercises or sports. Additionally, the proposed approach of real-time sensing of activities and bio-signals using in-ear sensors can also be extended to other lifestyle activities such as monitoring cognitive state and well-being of people in office environments.

Effect of weight lifted on variations in repetition pattern: From our user studies, we observed that people lifting heavier weights seem to get tired and thereby increase the repetition period. This could also be a way to infer 'mistakes' as the quality of exercise repetitions are degraded as the exercise set progresses. Similarly, we observed that when lifting heavier weights, individuals are prone to make comparatively more head movements during certain upper-body exercises (e.g., *lateral raise*) and the earable accelerometer is able to capture the exercise repetition peaks more clearly in such cases.

Exercises performed simultaneously: In the user studies conducted at our campus gym, we made all the subjects to start exercising at the exact same time. Anecdotally, we found that while there are multiple people (in pairs or small groups of 3 or 4) performing the *same* free-weights exercises concurrently, they do not necessarily start at the "exact" same time, thus simplifying the association problem. However, we believe that this may not be the case for group exercise classes and people's exercising patterns would be more time synchronized during such sessions.

Possible Additional Ways to Improve Performance: While incorporating additional sensor modalities is one way to improve the system performance, it may negatively impact the practicality of the approach (e.g., in terms of cost, form factor or deployment constraints). Therefore, in order to enhance the accuracy of associating the correct user and equipment pairs, a more practical approach is to devise additional features from existing sensor modalities. Capturing the 'range of motion' of the exercising equipment based on the inertial sensors (e.g., based on features such as the variance of the gyroscope) could be a potential feature that might aid in differentiating concurrent users' sensor streams. We suppose that individuals would exhibit a difference in their exercising motions (e.g., difference in the ways how arms are extended during a biceps curls exercise). For example, in cases where multiple individuals are concurrently performing the 'same' exercise, using the 3D trajectory of the exercising equipment as a feature in the association logic may have a better discriminatory power compared to temporal features which would be indistinguishable in cases where people are exercising in perfect time synchrony. Similarly, by combining features extracted from the ear-worn accelerometer sensor might help to identify additional head movements that may vary across individuals exercising concurrently.

4.7 Experiences and Lessons Learned

While experiments and user studies described as part of the research outlined in this Chapter were also conducted in a real gym (similar to as in Chapter 3), I found these particular user studies to be more cumbersome and involving new challenges and experiences. These challenges pertain to the added difficulty involved in conducting experiments with multiple concurrent participants. Additional care and attention were required to conduct all the experiments in a clean manner and to synchronize the study procedures across multiple individuals. I briefly describe below some of the lessons learned from this work.

- *Careful managing of multiple devices during the user study*: As part of these studies, we used a dedicated smartphone each (with our data collection app) to record the data from the earable and equipment sensor used by one individual. As such, mainly for the sessions involving more than two concurrent users, it needed special care and attention in handling say, four smartphone apps at the same time and in ensuring time synchronization. We had to discard some of the initial data recorded either because the clocks were not synchronized across all phones or when the individuals did not start exercising simultaneously.

- *Appropriate fitting of the earbud matters*: In the user study, there were instances (with 3 subjects) when the earbud fell off from the ear during an exercise set and all the users had to repeat the specific set. This was due to the inappropriate fitting of the earbud for certain subjects. This could be more problematic in our future scenarios where we intend to utilize physiological sensor data recorded from the earbud. Such physiological signals may not even get recorded properly if the earbud is not properly fitted.

4.8 Acknowledgments

I would like to specially thank Nokia Bell Labs and the research team behind the eSense platform for their generous contribution of earable devices, which I utilized for my research. I would also like to thank Tuan Tran for helping me with parts of the system implementation. I'm also extremely thankful to SMU gym staff for their help and cooperation during multiple phases of data collection and also to all the participants who willingly participated in the user study.

Chapter 5

Literature Review

Mobile phone sensing has emerged as a paradigm catering to multiple sectors such as healthcare, social networks, safety, environmental monitoring, transportation and retail. Wearable sensing has simultaneously evolved as a technology that enables human activity recognition at a finer granularity [33]. There is also an alternate body of research that focuses on leveraging only infrastructural sensors such as WiFi, video cameras or other IoT sensors deployed in the environment to capture detailed insights on individual's daily life activities. Activities in common daily life are often executed as a sequence of tasks each involving a specific gesture or interaction with daily artifacts. As gesture and fine-grained activity recognition lies at the foundation of all my contributions, I will first discuss some of the key works on pervasive sensing for gesture and interaction recognition. In this dissertation, my core contributions are in sensing gestures and various aspects of two common daily life activities – (a) *shopping* and (b) *exercising*. Therefore, I will present works which focus largely on monitoring shopping and exercise activities using disparate sensing modalities and compare our proposed approaches against those. I also present a summary of prior works in the behavioral literature on individual's exercise adherence and dropout behavior as well as digital tools proposed to motivate and support gym activity of individuals.

5.1 Gesture Recognition in Various Daily Lifestyle Activities

The feasibility of mobile sensing for human activity recognition has been well explored in literature [66, 135]. Human activity and gesture recognition is the core technology that enables a numerous number of applications in diverse areas, such as health care monitoring [26, 69], sleep monitoring [27, 72, 80], elderly care [58] and fitness tracking [34, 51]. There are also other class of applications which require gesture tracking in even finer granularity such as for touchscreen-based gesture input [16], pointing based interaction [90] and handwriting recognition [9]. Park et al. [85] proposed an adaptive threshold based segmentation technique for recognition of custom gestures during immersive lifestyle activities. Several works [10, 108, 122] demonstrate the detection of drinking and eating gestures using sensors from body-worn devices, RisQ system [84] depict how sensor data collected from an inertial measurement unit can be used to detect smoking gestures. While Blank et al. [20] studied the classification of Table Tennis strokes by instrumenting a racket with accelerometer and gyroscope sensors, Tran et al [130] proposed a technique for early recognition of table tennis gestures.

5.2 In-Store Shopping Activity Recognition

Unobtrusive monitoring of shopper's behavior inside physical retail stores presents interest for both the academic society and also for the marketing and private sector. There are also numerous case studies [47, 112, 14] on shopper/mall-level shopping behaviors which are typically confined to specific stores or demographics of shoppers. Several works [136, 99, 106] have explored the use of mobile or wearable sensors in unobtrusively monitoring the in-store shopper behavior. Our work utilizes a combination of such mobile and wearable sensing to uncover deeper insights into a shopper's in-store behavior. Unlike community-based personalized activity models [67], our work attempts to infer shopper behavior in a generalized setting where no shopper-specific training data is available. To the best of our knowledge, our work is among the first to utilize a mobile phone and smartwatch concurrently to infer item-level interactions of shoppers inside stores. The proposed approach does not pose any privacy concerns, is agnostic to user demographics and also does not require shopping history of users to infer shopping experience.

Table 5.1 provides a critical evaluation of the pros and cons of the proposed approach, *IRIS* against other alternative approaches in the literature for shopping activity recognition.

5.2.1 Mobile, Wearable and IoT Sensor-based Shopping Activity Monitoring

The interesting problem of studying the shopping time in stores is presented in [136], where a phone-based shopping tracker uses motif groups to identify movement trajectories and transforms the problem of monitoring shopping time as a classification problem. ThirdEye [99], uses image, inertial sensor, and WiFi data crowd-sourced from shoppers wearing smart glasses to track the physical browsing of shoppers. Sen et.al [106] proposes a person-independent activity recognition technique, CROSDAC, which uses smartphone based sensor (accelerometer, compass) data and WiFi, to identify the shopping intent of users. Lee et al. [68] presents an automated computing framework using smartphones designed to provide comprehensive understanding of customer behavior. Kanda et al. [52] uses a sensor network based on laser finders to predict people's shopping behaviours by clustering their accumulated trajectories.

Table 5.1: Comparison of various approaches for shopping activity recognition

Reference	Devices/Infrastructure Used	Approach	Insights Obtained	Advantages	Limitations
[92] Popa et al., 2010	CCTV	Video analysis	Model kinds of shoppers (goal-oriented, fun-shopper, etc.)	Personalized monitoring of shopping activities	Privacy issues
[136] You et al., 2011	Smartphone + WiFi AP	WiFi CSI analysis	Shopping vs Non-shopping time	Unobtrusive tracking of shopper trajectory	Does not identify fine-grained shopping activities
[68] Lee et al., 2013	Smartphone + WiFi AP	WiFi Based Trajectory	Stay detection, store recognition	Unobtrusive monitoring of shopper movement	Does not identify fine-grained shopping activities
[99] Rallapalli et al., 2014	Smartphone + Smartglass	Smartphone based location identification, smartglass based gaze detection	Shopping behavior (dwelling, gazing, reaching out), product layout	Requires no input from the store or the user	Privacy issues, smartglasses not widely adopted
[106] Sen et al., 2014	Smartphone + WiFi AP	Smartphone inertial sensors + WiFi	Shopping intention (buying intention, no buying intention, confused)	Handles user diversity	Does not identify fine-grained shopping activities
[138] Zeng et al., 2015	Smartphone + WiFi AP	WiFi CSI information	Walking behavior outside and inside a store	Cost-effective, privacy-, preserving, unobtrusive	Does not identify shopping gestures or item picked, does not work in multi-user environments
IRIS, 2015	Smartphone + Smartwatch	Smartwatch based gesture recognition & smartphone based location inference	Shopping gestures (pick, put-back, put-in-trolley & episode-level attributes (hurried, familiar)	Monitors fine-grained shopping gestures & shopper profile	Does not identify exact item picked
[110] Shangguan et al., 2015	RFID reader + tags	RFID phase change	Detect popular item category, hot items & correlated items	Unobtrusive monitoring of shopper browsing behavior & item of interest	Does not obtain individual-level shopper profile, RF-sensing prone to environmental changes
[107] Sen et al., 2018	Smartphone + BLE beacons	Smartwatch based gesture recognition and BLE based localization	Detects pick gesture, shelf & rack-level location of item interacted	Identifies all items interacted	RF-sensing prone to environmental changes
[139] Zang et al., 2019	RFID reader + tags + smartwatch	Correlation between inertial signals and RFID signals	Identifies user-item, item-item & user-user relations	Identifies items interacted without knowing product layout	RF-sensing prone to environmental changes

5.2.2 Infrastructural Sensor-based Shopping Activity Monitoring

Researchers have explored the use of video-based sensing for understanding various activities of shoppers in physical retail stores. Popa et al. [91] proposed a Kinect-based system for assessing shopping related actions. Based on the silhouette data from Kinect, they studied various in-store interactions such as whether the shopper is picking an item, trying on an item or interacting with the shopping cart. Liciotti et al. [70] also proposed an automated, integrated system that infers shopper behaviour from an RGB-D camera system. Another work [92] utilized the video feed from surveillance system to first identify the path taken by the shoppers inside a store and derived features to infer shopper's buying interest and the opportunities for making sales. Trinh et al. [124] proposes a finite state machine based approach to infer hand-activities in video-based retail surveillance. While these works leveraged video data for understanding shopper's activities or behavior, Zhang et al. [140] utilized the video data to study the social influence on shopping by extracting features which affect shopping such as the frequency of touch interactions with items, the trajectory taken inside retail stores. A global optimization framework based on binary quadratic programming (BQP) that seamlessly integrates appearance, motion and complex interactions between hands in video-based retail surveillance is proposed in [123]. Despite the fact the video-based approach is more straightforward in directly tracking all the actions and interactions of the shoppers inside a store, it poses privacy concerns as well as would fail to monitor the all the activities of a shopper due to occlusions, which would happen especially in case of a small or a crowded store.

Additionally, researchers have explored the use of RF-based approaches in understanding shopping behavior of people. ShopMiner [110] is an RFID-based system to infer the aggregated shopper interaction patterns with specific items in a physical clothing store. They utilized the difference in phase readings of RFID tags attached to individual items when shoppers are looking at an item, picking up an item or turning over an item. This approach requires attaching RFID tags to every item and also does not capture individual-level shopper profile and item interactions. Zeng et. al [138] utilized the Channel State Information of WiFi signals to infer a shopper's locomotive state (walking vs. standing) and location within a store. Lee et al. [68] also employs a WiFi based approach to recognize the store and analyze shopper's trajectory within a shopping mall. There are also commercial companies such as Euclid Analytics [35] that rely on sensing WiFi transmissions from shopper's smartphone to capture and analyze their in-store movements. Although WiFi based approaches can assist in identifying customer specific in-store movements, it cannot identify finer gestures and activities such as whether shopper picked an item or added an item to the cart. In contrast to these effort, our proposed IRIS system does not require any specific infrastructure support and rely primarily on sensors from individual shopper's mobile and wearable devices to obtain fine-grained insights on in-store activities.

5.3 Exercise Activity Recognition

Given the increasing emphasis on physical health and fitness, there has been a rapid surge in the market for fitness devices, applications and solutions. In this section, I focus on describing the relevant works in the domain of "exercise-monitoring" using mobile, wearable, infrastructural sensors to obtain quantified insights into different facets of a person's exercise routine and compare our approach against those.

Table 5.2 provides a comparison of the performance, pros and cons of the proposed *W8-Scope* approach against other alternative approaches for gym exercises monitoring. Note here that: (a) the accuracy comparisons for different components (across different works in the literature) shows a fair comparison as the attributes (e.g, repetition count, exercise type) captured are somewhat similar, (b) columns with 'N/A' means that the corresponding works do not capture those insights and hence, have no evaluation provided. For example, automatically determining the "amount of weight lifted" during weight machine exercises is a novel capability shown in our *W8-Scope* approach, which no existing works have evaluated previously.

5.3.1 Mobile, Wearable and IoT Sensor-based Exercise Monitoring

Most of the existing works [141, 32, 111, 75] on exercise monitoring has focused on segmenting exercises, recognizing exercise types and counting the repetitions of the exercise performed using either wearable or machine-attached sensors. Chang et al. [25] were one of the first to propose a wearable solution for tracking the type and repetition count of free-weight exercises, using multiple accelerometer sensors attached to a user's workout glove and waist. Another personalized approach [61], utilizes multiple body-worn sensors, together with sensors attached to the dumbbells, to detect anomalies in performing bicep curl exercises. MyHealthAssistant [104] is a system for classifying gym exercises using accelerometers attached on the hand, arm and leg. RecoFit [78] is also a wearable system based on an arm-worn inertial sensor to segment exercise and non-exercise periods and to detect different weight training and strength training exercises. Similarly, MiLift [111] is a smartwatch-based system that performs automatic segmentation and tracking of both cardio and weightlifting workouts. Mortazavi et al. [79] presented an approach to determine the best single sensor axis on a smartwatch for recognizing and counting repetitions of free weight and body weight exercises. Zhou et al. [141] proposed a wearable fabric pressure sensor system that measures the muscle movement, action and repetition of four different leg machine exercises. Burnout [75] is a sensor-embedded clothing to estimate skeletal muscle fatigue during isometric and isotonic exercises. Pernek et al. [89] developed a hierarchical algorithm, based on data from a wearable accelerometer attached to the upper body of subjects, to recognize intensity

Table 5.2: Comparison of various gym exercise activity recognition approaches

Reference	Approach	Target Exercises	Repetition Segmentation / Counting Accuracy	Exercise Classification Accuracy	Weight Classification Accuracy	User Classification Accuracy	Advantages	Limitations
[25] Chang et al., 2007	Multiple on-body accelerometer sensors	9 dumbbell exercises	94.1%	85%	N/A	N/A	Personalized tracking	Obtrusive, no weight and user identification,
[78] Morris et al., 2014	Arm-worn inertial sensor	13 types of different exercises	93%	93.8%	N/A	N/A	Personalized tracking	Obtrusive, no weight and user identification,
[32] Ding et al., 2015	RFID tag attached to dumbbell	10 dumbbell exercises	94%	90%	N/A	N/A	Unobtrusive	No weight & user identification, RF sensing prone to environmental changes, does not work in multi-user environment
[111] Shen et al., 2018	Smartwatch sensors	10 machine weight exercises and 5 dumbbell exercises	1.12 reps (avg error)	89.58%	N/A	N/A	Personalized tracking	Obtrusive, no weight and user identification,
[97] Rabbi et al., 2018	Sensor attached to terminal handles of exercise machine	12 machine weight exercises	97.96%	99.08%	N/A	N/A	Unobtrusive, VR-based real-time feedback	No weight & user identification
W8-Scope, 2018	Sensor attached to weight stack of exercise machine	14 exercises across a multi-purpose machine and 6 other dedicated weight machines	97%	96.9%	97.5%	98.9 %	Unobtrusive, personalized monitoring, robust to longitudinal changes	Requires training data from each individual
[41] Guo et al., 2018	CSI information from WiFi	10 different exercises (free-weights, aerobic)	Not specified	93%	N/A	97%	Unobtrusive	No weight identification, works only in home environment
[56] Khurana et al., 2018	Camera-based system	5 popular exercises	1.7 reps (avg error)	93.6%	N/A	N/A	Unobtrusive	Privacy concerns, no weight & user identification, may not work in occluded spaces
[19] Bian et al., 2019	Body capacitance-based sensor	7 different exercises	91%	63%	N/A	N/A	Tracks arm and leg exercises	Obtrusive, no weight and user identification

of strength training exercises. MyoBuddy [46] presents an approach to distinguish between different amounts of barbell weights using EMG signals recorded from an arm-worn EMG band. There are also other emerging apps and wearables such as TrackMyFitness [129] and Atlas Wristband [2] that auto-detects exercises, records repetitions and tracks workout progress. Unlike all these approaches which require the user to have some body-worn device, we propose a novel form of wearable-free and non-intrusive monitoring of a class of gym exercises.

There is an alternate body of prior work that assesses exercise characteristics using devices or sensors attached to different parts of the exercise machine. Moller et al. [77] explored the use of a smartphone-based trainer for assessing quality breakdown of exercises performed on a balance board. FEMO [32] is a platform for monitoring dumbbell exercises using passive RFID tags attached to individual dumbbells. As FEMO tracks dumbbell movements using RF signals, its performance could be affected by interference arising from the movement of other individuals in a multi-user gym. Sundholm et al. [119] developed a pressure sensor mat that recognizes and counts repetitions of 10 common strength training exercises performed on a mat. More recently, the Jarvis system [97] utilizes multiple IoT sensors, attached to different moving parts of exercise machine to segment repetitions, recognize exercise type and provide feedback to the user through a VR headset. Closest in spirit to our work, Jarvis also uses wearable EMG sensors to incorporate muscle activation activity as part of the feedback. In contrast, our approach uses a single sensor device mounted on a novel location (the weight stack) to support novel capabilities such as identifying the user performing the exercise and the amount of weight lifted (besides exercise recognition); we also consider the challenge of evolving the classifiers over medium time-scales.

5.3.2 Infrastructural Sensor-based Exercise Monitoring

Prior work has explored the use of WiFi [134, 41] and infrastructure-driven video sensing [45, 42, 127] for exercise activity recognition. SEARE [134] is a recent work that uses a WiFi based system and CSI waveform-based features for distinguishing between 4 different exercises. Similarly, Guo et al. [41] also introduce an approach that uses CSI information from WiFi infrastructure to provide workout interpretation and identify individuals exercising in a shared space within a home/work environment. However, these WiFi based systems may not work in a multi-user gym environment and in non line-of-sight scenarios. Havens et al. [45] proposed a technique for image-based contour tracking of spine and shoulder of the subjects from videos of treadmill exercises. A recently proposed system, GymCam [56] leverages a single camera to track multiple people exercising simultaneously and recognize their exercise type and repetitions. However, this system does not distinguish between users and also do not track other aspects of exercising such as the weight lifted, mistakes made. While, a less invasive approach [50] explores the use of thermal-imaging and optical flow techniques to estimate energy expenditure during treadmill exercises. Several works have used the Microsoft Kinect sensor for pose estimation and for tracking upper and lower body during simple rehabilitation exercises. Gonzalez-Ortega et al. [39] developed a 3D vision-based system to track the trajectories of human body parts during psychomotor exercises. Similarly, Dao et al. [31] present a system for monitoring the kinematics of exercises performed by elderly people during rehabilitation exercises. Another approach [42] uses the Kinect sensor to quantify the performance of squatting exercise using model-based metrics. Velloso et al. [127] presents a comparison of wearable sensor and Kinect model-based approaches for qualitative recognition of weight lifting exercises. All of these vision-based methods pose privacy concerns and are also highly dependent on external environment, such as adequate light and line of sight conditions. In contrast, we propose an alternative approach that is simpler to deploy, cost-effective and more privacy-friendly.

5.4 Behavioral Literature on Exercise Adherence & Digital Interventions

In this section, I present prior works that have studied the exercise adherence and dropout patterns of individuals and also provide a review of existing digital tools and technologies that are proposed by researchers to sustain motivation of exercisers as well as provide quantified insights into the exercise routine.

5.4.1 Studies on Exercise Adherence and Dropout

Trost et al. [125] present a review of the earlier literature that provide evidence relating to the personal, social, and environmental factors associated with physical activity. Similarly, Berger et al. [17] describe the aspects of psychological well-being that are influenced by physical activity and the factors that influence exercise participation. Existing works [105, 132, 53] have investigated the adherence behavior of people in specific exercise programs/physical activities and have reported that several factors (such as social support, guidance from staff, tangible health benefits) influence individual's motivation to continue in the program. Certain works have focused on understanding both the adherence and dropout behavior of specific user groups such as older adults (age above 50) [116], only women [48], low income groups [133], from various exercise programs. The works that specifically studied gym-goers [29, 54], have focused solely on understanding the motives of people for joining or continuing at the gym and not clearly identified the reasons to dropout. Pridgeon et al. [95] conducted a small scale study where they interviewed 14 gym-goers about their experiences in maintaining and dropping out of gym. Zarotis et al. [137] studies age-specific reasons to dropout from gym for different category of users. Most of these studies are purely interview-based or survey-based and conducted on a smaller scale of users. In this work, we present a more systematic study and provide quantified insights based on the actual digitally captured traces of individual-level gym visits, identify the key reasons for dropout and characterize some features that seem to affect dropout propensity.

5.4.2 Techniques to Improve Exercise Behavior

Prior works in the behavioral and sports science literature have proposed several techniques such as providing entertainment at the gym [15, 11], giving incentives [117], interventions with information of peer's gym attendance [103, 28] to sustain motivation of individuals to continue exercising. Although, mechanisms such as incentives tended to improve behavior during the intervention, findings were mixed on whether the observed improvements were sustained after incentives were removed. Hence, further research is required to derive appropriate mechanisms that are more personalized and can keep individuals motivated to persist their gym activity.

5.4.3 Digital Tools to Support Gym Activity

In the recent years, several commercial mobile applications (e.g., Trackmyfitness [129], JEFIT [49]) and wearable devices (e.g., Apple Watch, Nike Fuelband) have spawned in the fitness space with the goal to digitally track and encourage physical activity among individuals. However, a review of such physical activity apps found that only 2% provided evidence-based guidelines for gym exercises training and people find it not helpful [59]. There are also other works in the literature [24, 86, 37] that have proposed technologies for motivating and digital training during physical activities. Some of these approaches are based on health behaviour-change theories exploring features for motivating people to exercise. Patel et al. [87] study the contextual influence of digital technologies' use and non-use while exercising in gym based on interviews and participant observation. More recently, Rubin et al. [101] study the adoption factors of wearable technology in health and fitness space, specifically from a South African consumer perspective and identified that individuals did not enjoy using on-body devices during physical activity. This is similar to our finding from the survey conducted with gym-goers.

5.5 Ear-worn Sensing for Activity Recognition

Most of the prior works have explored the use of microphones in ear-worn devices to capture chewing sounds [10] and eating episodes [18], not many has explored the use of inertial sensors on ear-worn devices for complex activity recognition. Atallah et al. [13] proposed using an ear-worn accelerometer for gait monitoring while exercising on a treadmill. Nirjon et al. [82] proposed the 'MusicalHeart' system which uses a sensor-equipped ear-worn device that monitors heart rate and provides music recommendation based on user's activity levels. Gil et al. [38] developed a prototype of an ear-worn device that can measure cardiovascular and sweat parameters during physical exercise. In recent preliminary efforts, researchers have explored the potential of in-ear sensing for robust step counting [94], head-motion tracking [36] and more interestingly, for monitoring breathing rate [102]. In-ear inertial-sensing based respiration rate monitoring is of specific interest to us because combining such inferences on an individual's breathing pattern during exercise activity along with the head movement patterns (captured by the earable) might help in improving the performance of the associating the earable with the exercising equipment. For example, it would be interesting to see (i) if the inhalation/exhalation times match with the exercise repetition dynamics and, (ii) if the breathing patterns change during certain exercise sets (e.g., people may breathe heavily when lifting heavy weights).

5.6 Unobtrusive Device-Free Gesture Recognition

Recently, a new class of gesture recognition systems have spawned that utilizes wireless signals to track humans and identify their gestures [96, 131, 8, 71]. E-eyes [131] utilizes the WiFi channel state information (CSI) values to recognize gestures such as showering, brushing teeth etc. WiSee [96] is a WiFi based whole home sensing and gesture recognition and leverages the minute Doppler shifts and multi-path distortions that occur with these wireless signals from human motion in the environment. It works up to when 3 users are in the same room and recognize coarse-grained gestures. Soli [71] is a miniature FMCW radar that detects touch-less gesture interactions (sub-millimeter motion) of hand and uses a universal set of gestures to control devices. WiGest [8] utilizes RSSI changes, requires no training and achieves fine-grained gestures recognition for control of a specific user mobile device. However, a key limitation of these WiFi based gesture recognition is that none of them can recognize gestures that are simultaneously performed by multiple individuals in the environment.

5.6.1 Simultaneous Gesture Recognition of Multiple Users

Here I discuss some of the recent works that have tackled the problem of recognizing gestures performed simultaneously by multiple individuals in an indoor environment (not pertaining necessarily to shopping or exercise activities) using device-free approaches. Works such as [40, 60] have done some preliminary work in using UWB radars for multiple moving person tracking. However, they do not perform any user identification and tracking of the gesture performed. Peng et al. [88] presents preliminary results on recognizing gestures of multiple users using range-Doppler information from an FMCW radar. Recently proposed WiMu [128] system is the first one to track simultaneous gestures of multiple users. In WiMu, the effects of simultaneous movements of multiple users on CSI values are mathematically modeled to first detect that some users have performed gestures simultaneously, then identify the start and end times of the gestures and generate virtual samples of various combinations of those gestures. The main advantage of the system is that they do not require the users to provide training samples for all possible gesture combinations. However, the system only identifies *predefined gestures* performed simultaneously by multiple users and do not determine which user performed which gesture. It also cannot recognize continuous gestures. In Chapter 4, we demonstrate a hybrid approach of combining data sensed from unobtrusive wearable devices and cheap IoT sensors attached to the exercise equipment for monitoring of exercises performed simultaneously by people in a gym.

Chapter 6

Conclusion and Future Directions

In this chapter, I conclude this dissertation by summarizing the main contributions and outlining some of the possible extended use cases of the proposed technologies and key future directions.

6.1 Summary of Contributions

In this dissertation, I demonstrate the potential of leveraging sensors available in individual's personal devices or the sensors in cheap IoT devices that can be attached to objects in the indoor environment or their combination to both accurately and unobtrusively infer fine-grained aspects of daily lifestyle activities of individuals.

IRIS **:** In Chapter 2, I described the *IRIS* platform that uses standard locomotive and gestural micro-activities as building blocks to define novel composite features that help classify different facets of a shopper's interaction/experience with individual items, as well as attributes of the overall shopping episode or the store. IRIS utilizes inertial sensors on personal devices such as a smartphone and a smartwatch to infer micro-gestural activities and latent behavior of individual consumer behavior inside a retail store. We make the following key contributions:

1. **Robust and Accurate Segmentation:** I develop a novel, hierarchical segmentation algorithm to accurately delineate the (start, end) times of different

item-level interaction gestures, and aisle vs. non-aisle movements, over the entire duration of a store visit. With experiments conducted with 25 shoppers across 50 real-life grocery shopping episodes, we show that this technique is both *robust* (any mis-classifications never cascade beyond the current aisle) and *accurate* (it identifies gesture start and end times with mean errors of only 4.2 seconds, and achieves an overall 92% item-level gesture recognition accuracy).

2. **Accurate Recognition of Item-Level Interaction & Gesture-based Shopping Activities:** We show that *IRIS* can identify a variety of locomotive gestures (especially the {pick, put-in-trolley, put-back} gestures mentioned before), by appropriately using inertial sensor (accelerometer & gyroscope) based features from a smartwatch and a smartphone. Using these gestures as building blocks, we also subsequently infer *item-level* interactions such as whether the shopper buys the item frequently or knows specifically what he wants, using novel high-level features. All these classifications yield accuracies of over 90%.

3. **Accurate Prediction of Episode Attributes:** As the highest level of inference, we also utilize aggregate features (the item-level interaction history, plus in the in-aisle and non-aisle movement history) to build classifiers to estimate *episode-level* attributes, (such as "was the shopper in hurry?, and "did the shopper find the items he wanted?"), achieving accuracies of over 92%.

***W8-Scope*:** Chapter 3 describes the *W8-Scope* system which provides an unobtrusive and low-cost way to gather fine-grained, individual-specific insights into the exercise routines (including *mistakes* made) on a common class of weight stack machines. The key contributions made in this work involve the following:

1. **Novel Sensing Mode and Sensor Location for Exercise Monitoring:** We propose the use of a simple device, combining a 3-axis accelerometer and 3-

axis magnetometer sensor, mounted rigidly to the top plate of a weight stack to obtain fine-grained insights about the different exercises being performed.

2. **'Weight Stack Sensor' as a Viable Discriminator of Exercise Characteristics:** Using a set of validation studies performed using a commonplace *multi-exercise* "cable pulley" weight machine, we develop a *multi-stage* pipeline (called *W8-Scope*), combining magnetic and motion features, to infer multiple novel facets of exercises.

3. **Real-world Demonstration of W8-Scope:** With real world (*in-the-wild*) studies with regular gym-goers at two separate gyms: (a) a *University* gym and (b) a *Community* gym (open to the public), across 50 subjects performing 14 different exercises with a wide range of weights over 103 distinct sessions in these two gyms, we show that *W8-Scope* can identify the weight used with an accuracy of 97.5%, identify the exercise performed with 96.9% accuracy, detect commonplace mistakes made while exercising with 96.7% accuracy and also distinguish the user performing the exercise with over 98.7% accuracy (for a class of weight exercises and with co-terminuous training/test data).

4. **Longitudinal Tracking & Incremental Learning:** By adopting incremental learning techniques (that utilizes only highly confident samples to continually update the classifiers), *W8-Scope* can also accurately track these various facets of exercise over longitudinal periods, in spite of the inherent within-user differences that occur in exercising behaviors. Utilizing such approaches, we achieve an overall performance improvement of 12%, resulting in an accuracy of 90.2% for classifying exercises and 87.4% in distinguishing users over medium time-periods (12-15 weeks).

Finally, Chapter 4 presents a system that tracks weight-based exercises performed by multiple concurrent users in a gym and enables real-time audio-based corrective feedback to each exercising individual. In this work, we make the following key contributions:

1. **Earables as a platform for capturing fine-grained exercising aspects:** We introduce the vision of using ear-worn devices as the preferred, mass-market wearable platform, for both (a) individualized, fine-grained monitoring of gym exercise activities, and (b) subsequent real-time, context-aware feedback on exercise dynamics.

2. **Novel, hybrid architecture for multi-user gym environments:** We propose a low-cost solution that utilizes a hybrid architecture combining earables plus smart object/IoT on exercise equipment together for "superior activity recognition". We develop a matching technique that leverages novel temporal and wavelet-based features and inexact bipartite matching techniques to identify which individual is working out with which equipment (i.e., the correct $\{$equipment, earable$\}$ pair).

3. **Real-world evaluation of proposed approach:** Using 680 sets of real-world exercise data obtained with multiple people exercising concurrently, we show that: (a) our matching technique can achieve an accuracy of 88%, 65% and 45% in identifying all the distinct pairs when N=$\{2,3,4\}$ people are simultaneously performing weight-based exercises, and (b) by combining inertial sensor data-based features from both earable and equipment sensor, we can accurately identify the exercise performed (among 9 distinct choices) with 85% accuracy.

6.2 Future Directions

In this dissertation, I described novel systems and solutions for pervasive application scenarios that leverage on the judicious use of sensors on personal devices in combination with emerging infrastructure-based IoT devices to capture detailed insights of human lifestyle. I believe that the techniques developed and lessons learnt while building these systems will have deeper impact on the design of future ubiquitous systems and applications. Below I describe some of the possible extensions and future research directions:

Extended Shopping Applications & Scenarios: Real-time determination of the specific item being selected by shoppers in a store can be used for other types of consumer-specific alerts. Consumers today can use their mobile devices to obtain instant information (e.g., customer reviews, product ratings or price comparisons) from online sources. At present, such information retrieval typically requires manual input– i.e., the shopper must either upload a picture or a product specification to the online service. Real-time wearable+IoT analytics offers the possibility of making such retrieval unobtrusive. For example, if the item picked up by the shopper turns out to have ingredients to which the shopper is allergic, a product alert application can proactively alert the shopper to such inadvertent selections. Similar, more accurate tracking of the numbers selected, for a specific item, might alert the shopper to possible promotions and deals that she may be unaware of. For example, if a particular brand of apples has a "3 for \$2" offer (with a unit price of \$2), a deal detective application can automatically alert the shopper about the promotional offer, if it detects that she has selected only 2 apples.

Additional Use-cases of IRIS Technologies: I believe that the gesture recognition technologies developed using the combination of sensor data from a smartphone and smartwatch can be extended to other application domains such as elderly care, smart manufacturing. For example, applications for monitoring the well-being of the elderly can utilize the *IRIS* technologies to automatically track gestures/activities (e.g., drinking water, regular in-take of medicines or other locomotive activities). Similarly, in a manufacturing setting, smartwatches worn by factory workers can be used to track their activities in a PCB-manufacturing/assembling environment. This will help in potential applications for early-tracking of mistakes made by workers in assembling the unit (e.g., certain tasks are meant to be done with one hand or sequentially and using both hands would be flagged as a potential violation).

Device-free Recognition of Multi-user Gestures & Behaviors: The techniques and solutions proposed in this dissertation are primarily based on the fusion of inertial sensor data from personal and infrastructural IoT devices. More recently, there has been a significant attention on wireless or device-free gesture recognition systems that utilizes Wi-Fi, sound signals, etc.–such approaches have the advantage that they do not require the individual to carry any electronic device, but instead rely on the signal variations (e.g., channel state information (CSI) from Wi-Fi) induced by human movements. However, a key limitation of the prior works on gesture recognition is that they are intended for scenarios involving only a single individual and such solutions fail to work when multiple users simultaneously perform some gestures. To overcome this limitation, I pose the question: *Given the deployment of multiple short-range radar devices in the indoor space, can we accurately identify that multiple individuals are present in the indoor space and are simultaneously performing some gestures/activities and also identify gestures of each individual using analytics techniques that utilize both the data from multiple short-range radars and from low-power wearable devices?*

In-ear Sensing of other Activities: As sensory earables are increasingly becoming popular, it opens up new opportunities and challenges in the space of personal-scale human sensing applications. For example, inertial sensors in earables could be used to build a *toothbrushing* application that would detect (based on micro head-movements) if all areas inside the mouth are brushed properly and provide real-time corrective audio-based feedback otherwise. Such an application would especially be useful for the kids. Most of the in-ear sensing applications are based on inferences derived from head-motion patterns. So, I ask *can we use such motion signatures to accurately track an individual's "gazing direction" and enable interactive applica-*

tions and how does the performance compare to that of other state-of-the-art gaze tracking techniques? I believe earables also has a huge potential in enabling several health sensing applications (e.g., monitoring breathing patterns).

Appendix

Survey Questionnaire Distributed in University Campus

1. Do you go to a gym?

   ○ Goes to SMU Gym

   ○ Goes to another gym

   ○ Used to go to gym, has stopped going now

2. If you have stopped going to a gym (or dropped out), what made you stop? Please rate each of the below reasons on a scale of 1 to 5 (1 for "Not at all Important" and 5 for "Extremely Important")

   - Don't see the benefits
   - Lack of social support
   - Lack of enjoyment
   - Lack of knowledge in using gym equipment
   - Lack of personal trainer
   - Lack of time
   - Initial overhead in getting to the gym
   - Fatigue from over-training

- Not meeting goals/expectations from the workout
- Medical reasons
- Prefer some other workout (e.g., running, yoga, zumba, martial arts, free hand exercise at home)
- Only dieting for now
- Going to another gym or fitness center

3. Rate the services that would be important for you when deciding to continue going to the gym. (Rate each of the below statement on a scale from 1 to 5 (1 for "Not at all Important" and 5 for "Extremely Important")

    - Having a friend to accompany you
    - More entertainment at the gym
    - Personal training recommendations
    - Availability of exercise classes (e.g., Yoga, martial arts, aerobics)
    - Provide more awareness about fitness
    - Provide nutritional tips and suggestions
    - More variety of exercise machines
    - Provide motivational tips to continue going to gym

4. What else would help you to continue going to the gym or to improve your experience at the gym. Please tell as much as you can about what would help you continue using the gym.

5. How do you self-rate your gym usage? Please select all options that apply to you.

    - ☐ Cardio Zone (Campus Green level)
    - ☐ Mezzanine level Green (Free weights) Zone

- ◻ Mezzanine level Blue (Functional Training) Zone
- ◻ Mezzanine level Red (Strength Conditioning) Zone

6. How often do you visit the Gym?
   - ○ Less than once a month
   - ○ Once a month
   - ○ Once in two weeks
   - ○ Once a week
   - ○ Twice a week
   - ○ Thrice a week
   - ○ More than three times a week

7. I've been going to a gym (any gym) regularly for the last ...
   - ○ Less than a month
   - ○ 1-3 months
   - ○ 4-6 months
   - ○ 7-11 months
   - ○ 1 year
   - ○ 2-4 years
   - ○ 4+ years

8. Rate your reasons for going to the gym (For each of the below statement, choose a scale point from 1 for "Not at all true for me" to 5 for "Completely True for me").
   - Enhance your athletic/sport performance
   - Train for muscle building/power lifting

- Lose weight
- Maintain my physical health and well-being
- Improve my body shape and appearance
- To manage stress and tension
- Medical reasons
- Meet people and socialize
- For fun and relaxation
- For the challenge and excitement of participation
- To be fit and stay healthy

9. Please indicate for each of the below mentioned exercises, how often you do it. Rate on the scales: 'On Each Visit', 'On Most Visits', 'Infrequently', 'Very Rarely', 'Never'.

    - Cardio exercises (on treadmill, elliptical, exercise bike)
    - Weight training using the weight machines (leg press, shoulder press etc.)
    - Weight training using free weights
    - Free-hand or mat exercises (push-ups, crunches, sit-ups etc.)
    - Circuit training (Intense combination of the above routines repeated multiple times)
    - Group Exercise Classes (Dance, Zumba, Yoga, Pilates, etc.)

10. Which are the machines that use during your gym workout sessions? Please select *ALL* options that apply to you.

    - ☐ Treadmill
    - ☐ Elliptical

☐ Exercise bikes

☐ Lateral Trainer

☐ Cable pulley machine/exercises

☐ Core trainer

☐ Rowing machine

☐ Ground Base Combo Decline (Arm exercises)

☐ Chest Press/Lateral Pull Down

☐ Deadlifts machine

☐ Shoulder/Upper-back exercise machine

☐ Chin up/Bench press machine

☐ Squats/Deadlift/Military Press machine

☐ Jump trainer

☐ Free-weights with dumbbells (please specify what exercises) ☐

☐ Free-weights with barbells and plates (please specify what exercises) ☐

☐ Floor/mat exercises (please specify what exercises) ☐

☐ Other machines/exercises (please specify) ☐

11. Would you prefer to use a fitness app/wearable that monitors your exercise form during your workout and provides personalized feedback on your form as well as suggesting corrective actions and alternative exercises (with the goal of making your exercise routine more effective and / or safer)?

   ○ Already using an app

   ○ Maybe (interested to use in the future)

○ No or Stopped using

12. If you are already using a fitness app, please mention the app name. Also list the key features in the App that you find useful/use.

13. If you are interested to use a fitness app in future, what would you want to mainly use it for ? Rank the below listed features (by scrolling and reordering the options) in your order of preference (i.e., based on what values to you most when deciding to use a fitness app).

    - Set personalized goals and exercise regimen
    - Automatically track all exercises performed and provide summary reports
    - Provide personalized feedback or suggestions (e.g., specific muscle groups that user need to train more)
    - Identify mistakes (e.g., incorrect use of gym equipment, incorrect body postures) made while exercising and provide corrective feedback
    - Provide nutritional tips
    - Teach you how to perform specific exercises

14. If you are not using (or stopped using) a fitness App, please specify your reasons. (Rate each of the below statement on a scale from 1 for ”Not at all true of me” to 5 for ”Extremely true of me”)

    - Do not want to use any device (phone, wearables)
    - Used apps did not meet expectations (too confusing, time consuming)
    - Provided suggestions were too generic
    - Hidden cost (expensive premium features)
    - Apps are intrusive (not comfortable with sharing data)

- Other (please specify) [ ]

15. How valuable would it be for you to have access to a personal trainer? Rate on a scale from 1 “Not at all valuable” to 5 “Extremely valuable”

16. What are the various things that you think a personal trainer can help you with? (Rate each of the below statement on a scale from 1 to 5 (1 for “Not at all Important” and 5 for “Extremely Important”)

    - Discuss your personal needs and assess fitness
    - Help to set short-term and long- term goals
    - Help with setting a personalized exercise regimen
    - Teach you how to do the exercises
    - Help with correcting form/posture
    - Provide motivation during exercising
    - Provide nutritional tips and meals plans
    - Other (please specify) [ ]

17. Please provide your SMU email address.

    [ ]

Survey Questionnaire Distributed to General Public in Singapore

This questionnaire was distributed in public gyms in Singapore and also hosted in Amazon Mechanical Turk. The questions in this survey were mostly similar to that distributed in the University campus. Therefore, for a succinct presentation, below I present only the additional questions that were included in this version of the survey.

1. Would you prefer to wear a wearable device (such as a smartwatch/smartband, VR head display) that monitors your exercise form during your workout and provides personalized feedback on your form as well as suggest corrective actions and alternative exercises (with the goal of making your exercise routine more effective and / or safer)?

   ○ Yes, I use a wearable device while exercising

   ○ No or Stopped using

   ○ May be (interested to use in the future)

2. If you are already using a wearable fitness tracker, please mention the wearable device name. Also list the key reasons why you use it or the key features that you find useful/use.

3. If you are not using (or stopped using) a Wearable fitness tracker, please specify your reasons. (Rate each of the below statement on a scale from 1 for ”Not at all true of me” to 5 for ”Extremely true of me”)

   - Do not want to use any device as it causes inconvenience while exercising.
   - Wearable devices that I tried were not useful (did not track any useful info).
   - Did not provide any personalized recommendation based on data that was tracked.
   - Not comfortable with tracking and sharing of exercise data).

4. Imagine that a futuristic technology can perform fine-grained monitoring and tracking of your exercises performed in a Gym and provide you with personalized feedback on your form as well as suggest corrective actions and alter-

native exercises (with the goal of making your exercise routine more effective and / or safer).

The technology could be either (i) a wearable-based solution with on-body sensors (e.g., a wristband, smart clothing) or (ii) solution based on simple IoT sensors (of small form-factor) attached to the exercise machines itself (i.e., no instrumentation on body). Which approach would you prefer and why?

- ○ Wearable approach (*Please specify why you prefer this approach over the other.)

  ____

- ○ Machine-based sensor approach (*Please specify why you prefer this approach over the other.)

  ____

5. Please select your gender.

   - ○ Male
   - ○ Female
   - ○ Do not wish to specify

6. Please select your age group.

   - ○ 18-20
   - ○ 21-25
   - ○ 26-30
   - ○ 31-35
   - ○ 36-40
   - ○ 41-45
   - ○ 46-50

○ 51-55

○ 56-60

○ 61-65

○ Above 65

7. What is your current employment status?

○ Employed Full time

○ Employed Part time

○ Self-employed

○ Student

○ Homemaker

○ Retired

Bibliography

[1] Amazon go. https://www.amazon.com/b?node=16008589011; [Online Last Accessed: May 2019].

[2] Atlas wearables, 2018. https://atlaswearables.com; [Online Last Accessed: December 2018].

[3] Infineon demo position2go board, 2018. https://www.infineon.com/cms/en/product/evaluation-boards/demo-position2go/; [Online Last Accessed: May 2019].

[4] Beacon–synergizing the best of both worlds, 2019. https://www.jcdecaux.com.sg/news-and-press-releases/beacon-synergizing-best-both-worlds/; [Online Last Accessed: November 2019].

[5] Statista, 2019. https://www.statista.com/topics/1556/wearable-technology/; [Online Last Accessed: May 2019].

[6] Top proximity marketing apps for 2019, 2019. https://www.mobileappdaily.com/top-proximity-marketing-apps/; [Online Last Accessed: November 2019].

[7] Ccs insight wearables forecast, October 2018. https://is.gd/oUJFey; [Online Last Accessed: May 2019].

[8] H. Abdelnasser, M. Youssef, and K. A. Harras. Wigest: A ubiquitous wifi-based gesture recognition system. In *2015 IEEE Conference on Computer Communications (INFOCOM)*, pages 1472–1480. IEEE, 2015.

[9] S. Agrawal, I. Constandache, S. Gaonkar, R. Roy Choudhury, K. Caves, and F. DeRuyter. Using mobile phones to write in air. In *Proceedings of the 9th international conference on Mobile systems, applications, and services*, pages 15–28. ACM, 2011.

[10] O. Amft, H. Junker, and G. Troster. Detection of eating and drinking arm gestures using inertial body-worn sensors. In *Ninth IEEE International Symposium on Wearable Computers (ISWC'05)*, pages 160–163. IEEE, 2005.

[11] J. J. Annesi. Effects of music, television, and a combination entertainment system on distraction, exercise adherence, and physical output in adults. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 33(3):193, 2001.

[12] W. Applebaum. Studying customer behavior in retail stores. *The Journal of Marketing*, pages 172–178, 1951.

[13] L. Atallah, A. Wiik, G. G. Jones, B. Lo, J. P. Cobb, A. Amis, and G.-Z. Yang. Validation of an ear-worn sensor for gait monitoring using a force-plate instrumented treadmill. *Gait & posture*, 35(4):674–676, 2012.

[14] M. Avello, D. Gavilán, C. Abril, and R. Manzano. Experiential shopping at the mall: influence on consumer behaviour. *China-USA Business Review*, 10(1):16–24, 2011.

[15] M. J. Barwood, N. J. Weston, R. Thelwell, and J. Page. A motivational music and video intervention improves high-intensity exercise performance. *Journal of sports science & medicine*, 8(3):435, 2009.

[16] O. Bau and W. E. Mackay. Octopocus: a dynamic guide for learning gesture-based command sets. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, pages 37–46. ACM, 2008.

[17] B. Berger. Pargman, d., & weinberg, rs (2002). *Foundations of exercises psychology*, 2002.

[18] S. Bi, T. Wang, N. Tobias, J. Nordrum, S. Wang, G. Halvorsen, S. Sen, R. Peterson, K. Odame, K. Caine, et al. Auracle: Detecting eating episodes with an ear-mounted sensor. *Proc. of the ACM IMWUT*, 2(3):92, 2018.

[19] S. Bian, V. Rey, P. Hevesi, and P. Lukowicz. Passive capacitive based approach for full body gym workout recognition and counting. In *Proc. of PerCom'19*.

[20] P. Blank, J. Hoßbach, D. Schuldhaus, and B. M. Eskofier. Sensor-based stroke detection and stroke type classification in table tennis. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pages 93–100. ACM, 2015.

[21] BodyBuilding.com. Cable exercises & exercise guides, 2018. `https://www.bodybuilding.com/exercises/equipment/cable`; [Online Last Accessed: July 2018].

[22] N. E. Breslow. Analysis of survival data under the proportional hazards model. *International Statistical Review/Revue Internationale de Statistique*, pages 45–57, 1975.

[23] J. Burnham. Exercise is medicine: health benefits of regular physical activity. *The Journal of the Louisiana State Medical Society: official organ of the Louisiana State Medical Society*, 150(7):319–323, 1998.

[24] F. Buttussi, L. Chittaro, and D. Nadalutti. Bringing mobile guides and fitness activities together: a solution based on an embodied virtual trainer. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*, pages 29–36. ACM, 2006.

[25] K.-H. Chang, M. Y. Chen, and J. Canny. Tracking free-weight exercises. In *International Conference on Ubiquitous Computing*, pages 19–37. Springer, 2007.

[26] L. Y. Chen, B. C.-K. Tee, A. L. Chortos, G. Schwartz, V. Tse, D. J. Lipomi, H.-S. P. Wong, M. V. McConnell, and Z. Bao. Continuous wireless pressure monitoring and mapping with ultra-small passive sensors for health monitoring and critical care. *Nature communications*, 5:5028, 2014.

[27] Z. Chen, M. Lin, F. Chen, N. D. Lane, G. Cardone, R. Wang, T. Li, Y. Chen, T. Choudhury, and A. T. Campbell. Unobtrusive sleep monitoring using smartphones. In *Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare*, pages 145–152. ICST (Institute for Computer Sciences, Social-Informatics and ..., 2013.

[28] S. Condliffe, E. Işgın, and B. Fitzgerald. Get thee to the gym! a field experiment on improving exercise habits. *Journal of behavioral and experimental economics*, 70:23–32, 2017.

[29] N. Crossley. In the gym: Motives, meaning and moral careers. *Body & Society*, 12(3):23–50, 2006.

[30] S. J. Czaja, N. Charness, A. D. Fisk, C. Hertzog, S. N. Nair, W. A. Rogers, and J. Sharit. Factors predicting the use of technology: findings from the center for research and education on aging and technology enhancement (create). *Psychology and aging*, 21(2):333, 2006.

[31] T. Dao, H. Tannous, P. Pouletaut, D. Gamet, D. Istrate, and M. H. B. Tho. Interactive and connected rehabilitation systems for e-health. *Irbm*, 37(5-6):289–296, 2016.

[32] H. Ding, L. Shangguan, Z. Yang, J. Han, Z. Zhou, P. Yang, W. Xi, and J. Zhao. Femo: A platform for free-weight exercise monitoring with rfids. In *Proc. of ACM SenSys'15*, pages 141–154. ACM, 2015.

[33] O. DL and M. AL. A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys and Tutorials*, 15(3):1192–1209, 2013.

[34] C. Dondzila and D. Garner. Comparative accuracy of fitness tracking modalities in quantifying energy expenditure. *Journal of medical engineering & technology*, 40(6):325–329, 2016.

[35] Euclid Analytics. Euclid analytics. `https://geteuclid.com/`; [Online Last Accessed: April 2019].

[36] A. Ferlini, A. Montanari, M. Cecilia, and R. Harle. Head motion tracking through in-ear wearables. In *Proc. of EarComp'19*, 2019.

[37] T. Fritz, E. M. Huang, G. C. Murphy, and T. Zimmermann. Persuasive technology in the real world: a study of long-term use of activity sensing devices for fitness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 487–496. ACM, 2014.

[38] B. Gil, S. Anastasova, and G. Z. Yang. A smart wireless ear-worn device for cardiovascular and sweat parameter monitoring during physical exercise: Design and performance results. *Sensors*, 19(7):1616, 2019.

[39] D. González-Ortega, F. Díaz-Pernas, M. Martínez-Zarzuela, and M. Antón-Rodríguez. A kinect-based system for cognitive rehabilitation exercises monitoring. *Computer methods and programs in biomedicine*, 113(2):620–631, 2014.

[40] B. Gulmezoglu, M. B. Guldogan, and S. Gezici. Multiperson tracking with a network of ultrawideband radar sensors based on gaussian mixture phd filters. *IEEE Sensors Journal*, 15(4):2227–2237, 2015.

[41] X. Guo, J. Liu, C. Shi, H. Liu, Y. Chen, and M. C. Chuah. Device-free personalized fitness assistant using wifi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(4):165, 2018.

[42] S. Hagler, H. B. Jimison, R. Bajcsy, and M. Pavel. Quantification of human movement for assessment in automated exercise coaching. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pages 5836–5839. IEEE, 2014.

[43] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[44] M. A. Hall. Correlation-based feature subset selection for machine learning. *Thesis submitted in partial fulfillment of the requirements of the degree of Doctor of Philosophy at the University of Waikato*, 1998.

[45] T. C. Havens, G. L. Alexander, C. Abbott, J. M. Keller, M. Skubic, and M. Rantz. Contour tracking of human exercises. In *Computational Intelligence for Visual Intelligence, 2009. CIVI'09. IEEE Workshop on*, pages 22–28. IEEE, 2009.

[46] B.-J. Ho, R. Liu, H.-Y. Tseng, and M. Srivastava. Myobuddy: Detecting barbell weight using electromyogram sensors. In *Proceedings of the 1st Workshop on Digital Biomarkers*, pages 27–32. ACM, 2017.

[47] H. Hu and C. Jasper. A qualitative study of mall shopping behaviors of mature consumers. *Journal of Shopping Center Research*, 14(1):17–38, 2007.

[48] J. L. Huberty, L. B. Ransdell, C. Sidman, J. A. Flohr, B. Shultz, O. Grosshans, and L. Durrant. Explaining long-term exercise adherence in women who complete a structured exercise program. *Research Quarterly for exercise and Sport*, 79(3):374–384, 2008.

[49] Jefit Inc. [US]. Jefit. `https://www.jefit.com/`; [Online Last Accessed: May 2019].

[50] M. M. Jensen, M. K. Poulsen, T. Alldieck, R. G. Larsen, R. Gade, T. B. Moeslund, and J. Franch. Estimation of energy expenditure during treadmill exercise via thermal imaging. *Medicine and science in sports and exercise*, 48(12):2571–2579, 2016.

[51] K. Kaewkannate and S. Kim. A comparison of wearable fitness devices. *BMC public health*, 16(1):433, 2016.

[52] T. Kanda, D. F. Glas, M. Shiomi, H. Ishiguro, and N. Hagita. Who will be the customer?: a social robot that anticipates people's behavior from their trajectories. In *Proc. of UbiComp'08*.

[53] L. Karvitz. Exercise motivation: What starts and keeps people exercising. *University of New Mexico*, 2011.

[54] N. Kaushal and R. E. Rhodes. Exercise habit formation in new gym members: a longitudinal study. *Journal of Behavioral Medicine*, 38(4):652–663, 2015.

[55] Z. Y. Kerr, C. L. Collins, and R. Dawn Comstock. Epidemiology of weight training-related injuries presenting to united states emergency departments, 1990 to 2007. *The American Journal of Sports Medicine*, 38(4):765–771, 2010.

[56] R. Khurana, K. Ahuja, Z. Yu, J. Mankoff, C. Harrison, and M. Goel. Gymcam: Detecting, recognizing and tracking simultaneous exercises in unconstrained scenes. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(4):185, 2018.

[57] A. C. King, E. B. Hekler, L. A. Grieco, S. J. Winter, J. L. Sheats, M. P. Buman, B. Banerjee, T. N. Robinson, and J. Cirimele. Effects of three motivationally targeted mobile device applications on initial physical activity and sedentary behavior change in midlife and older adults: a randomized trial. *PloS one*, 11(6):e0156370, 2016.

[58] D. E. King, E. Matheson, S. Chirina, A. Shankar, and J. Broman-Fulks. The status of baby boomers' health in the united states: the healthiest generation? *JAMA internal medicine*, 173(5):385–386, 2013.

[59] E. Knight, M. I. Stuckey, H. Prapavessis, and R. J. Petrella. Public health guidelines for physical activity: is there an app for that? a review of android and apple app stores. *JMIR mHealth and uHealth*, 3(2), 2015.

[60] D. Kocur, J. Fortes, and M. Švecová. Multiple moving person tracking by uwb sensors: the effect of mutual shielding persons and methods reducing its impacts. *EURASIP Journal on Wireless Communications and Networking*, 2017(1):68, 2017.

[61] Y. Kowsar, M. Moshtaghi, E. Velloso, L. Kulik, and C. Leckie. Detecting unseen anomalies in weight training exercises. In *Proc. of ACM OzCHI'16*, pages 517–526. ACM, 2016.

[62] P. Krebs and D. T. Duncan. Health app use among us mobile phone owners: a national survey. *JMIR mHealth and uHealth*, 3(4), 2015.

[63] J. Krockel and F. Bodendorf. Intelligent processing of video streams for visual customer behavior analysis. In *Proc. of ICONS'12*.

[64] H. W. Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.

[65] S. Y. Lam. The effects of store environment on shopping behaviors: A critical review. *ACR North American Advances*, 2001.

[66] N. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. Campbell. A survey of mobile phone sensing. *Comm. Mag.*, 48(9), Sept. 2010.

[67] N. Lane, Y. Xu, H. Lu, S. Hu, T. Choudhury, A. Campbell, and F. Zhao. Enabling large-scale human activity inference on smartphones using community similarity networks (csn). In *Proc. of UbiComp '11*.

[68] S. Lee, C. Min, C. Yoo, and J. Song. Understanding customer malling behavior in an urban shopping mall using smartphones. In *Proc. of MSCSS'13 colocated with UbiComp'13*.

[69] C. Li, V. M. Lubecke, O. Boric-Lubecke, and J. Lin. A review on recent advances in doppler radar sensors for noncontact healthcare monitoring. *IEEE Transactions on microwave theory and techniques*, 61(5):2046–2060, 2013.

[70] D. Liciotti, M. Contigiani, E. Frontoni, A. Mancini, P. Zingaretti, and V. Placidi. Shopper analytics: a customer activity recognition system using a distributed RGB-D camera network. *CoRR*, abs/1508.06853, 2015.

[71] J. Lien, N. Gillian, M. E. Karagozler, P. Amihood, C. Schwesig, E. Olson, H. Raja, and I. Poupyrev. Soli: Ubiquitous gesture sensing with millimeter wave radar. *ACM Transactions on Graphics (TOG)*, 35(4):142, 2016.

[72] J.-K. Min, A. Doryab, J. Wiese, S. Amini, J. Zimmerman, and J. I. Hong. Toss'n'turn: smartphone as sleep and sleep quality detector. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 477–486. ACM, 2014.

[73] M. Mirtchouk, C. Merck, and S. Kleinberg. Automated estimation of food type and amount consumed from body-worn audio and motion sensors. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 451–462. ACM, 2016.

[74] W. W. Moe. Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. *Journal of consumer psychology*, 13(1-2):29–39, 2003.

[75] F. Mokaya, R. Lucas, H. Y. Noh, and P. Zhang. Burnout: a wearable system for unobtrusive skeletal muscle fatigue estimation. In *Proc. of IPSN'16*, page 8. IEEE Press, 2016.

[76] J. S. Mollee, A. Middelweerd, R. L. Kurvers, and M. C. Klein. What technological features are used in smartphone apps that promote physical activity? a review and content analysis. *Personal and Ubiquitous Computing*, 21(4):633–643, 2017.

[77] A. Möller, L. Roalter, S. Diewald, J. Scherr, M. Kranz, N. Hammerla, P. Olivier, and T. Plötz. Gymskill: A personal trainer for physical exercises. In *Proc. of IEEE PerCom'12*, pages 213–220. IEEE, 2012.

[78] D. Morris, T. S. Saponas, A. Guillory, and I. Kelner. Recofit: using a wearable sensor to find, recognize, and count repetitive exercises. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3225–3234. ACM, 2014.

[79] B. J. Mortazavi, M. Pourhomayoun, G. Alsheikh, N. Alshurafa, S. I. Lee, and M. Sarrafzadeh. Determining the single best axis for exercise repetition recognition and counting on smartwatches. In *Proc. of BSN'14*, pages 33–38. IEEE, 2014.

[80] R. Nandakumar, S. Gollakota, and N. Watson. Contactless sleep apnea detection on smartphones. In *Proceedings of the 13th annual international conference on mobile systems, applications, and services*, pages 45–57. ACM, 2015.

[81] V. Natale, M. Drejak, A. Erbacci, L. Tonetti, M. Fabbri, and M. Martoni. Monitoring sleep with a smartphone accelerometer. *Sleep and Biological Rhythms*, 10(4):287–292, 2012.

[82] S. Nirjon, R. F. Dickerson, Q. Li, P. Asare, J. A. Stankovic, D. Hong, B. Zhang, X. Jiang, G. Shen, and F. Zhao. Musicalheart: A hearty way of listening to music. In *SenSys'12*, 2012.

[83] A. C. of Sports Medicine et al. American college of sports medicine position stand. progression models in resistance training for healthy adults. *Medicine and science in sports and exercise*, 41(3):687, 2009.

[84] A. Parate, M.-C. Chiu, C. Chadowitz, D. Ganesan, and E. Kalogerakis. Risq: Recognizing smoking gestures with inertial sensors on a wristband. In *Proc. of MobiSys'14*.

[85] T. Park, J. Lee, I. Hwang, C. Yoo, L. Nachman, and J. Song. E-gesture: a collaborative architecture for energy-efficient gesture recognition with hand-worn sensor and mobile devices. In *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems*, pages 260–273. ACM, 2011.

[86] T. Park, U. Lee, B. Lee, H. Lee, S. Son, S. Song, and J. Song. Exersync: interpersonal synchrony in social exergames. In *Proceedings of the 2013 conference on Computer supported cooperative work companion*, pages 27–30. ACM, 2013.

[87] M. Patel and A. A. O'Kane. Contextual influences on the use and non-use of digital technology while exercising at the gym. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2923–2932. ACM, 2015.

[88] Z. Peng, C. Li, J.-M. Muñoz-Ferreras, and R. Gómez-García. An fmcw radar sensor for human gesture recognition in the presence of multiple targets. In *2017 First IEEE MTT-S International Microwave Bio Conference (IMBIOC)*, pages 1–3. IEEE, 2017.

[89] I. Pernek, G. Kurillo, G. Stiglic, and R. Bajcsy. Recognizing the intensity of strength training exercises with wearable sensors. *Journal of biomedical informatics*, 58:145–155, 2015.

[90] K. Pietroszek, L. Tahai, J. R. Wallace, and E. Lank. Watchcasting: Freehand 3d interaction with off-the-shelf smartwatch. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 172–175. IEEE, 2017.

[91] M. Popa, A. K. Koc, L. J. Rothkrantz, C. Shan, and P. Wiggers. Kinect sensing of shopping related actions. In *International joint conference on ambient intelligence*, pages 91–100. Springer, 2011.

[92] M. Popa, L. Rothkrantz, Z. Yang, P. Wiggers, R. Braspenning, and C. Shan. Analysis of shopping behavior based on surveillance system. In *Proc. of SMC'10*.

[93] R. Poppe. A survey on vision-based human action recognition. *Image and vision computing*, 28(6):976–990, 2010.

[94] J. Prakash, Z. Yang, Y.-L. Wei, and R. R. Choudhury. Stear: Robust step counting from earables. In *Proc. of EarComp'19*, 2019.

[95] L. Pridgeon and S. Grogan. Understanding exercise adherence and dropout: an interpretative phenomenological analysis of men and women's accounts of gym attendance and non-attendance. *Qualitative research in sport, Exercise and Health*, 4(3):382–399, 2012.

[96] Q. Pu, S. Gupta, S. Gollakota, and S. Patel. Whole-home gesture recognition using wireless signals. In *Proceedings of the 19th annual international conference on Mobile computing & networking*, pages 27–38. ACM, 2013.

[97] M. F. Rabbi, T. Park, B. Fang, M. Zhang, and Y. Lee. *When Virtual Reality Meets IoT in the Gym: Enabling Immersive and Interactive Machine Exercise*. ACM, 2018.

[98] M. Radhakrishnan, S. Sen, S. Vigneshwaran, A. Misra, and R. Balan. Iot+ small data: Transforming in-store shopping analytics & services. In *2016 8th international conference on communication systems and networks (COMSNETS)*, pages 1–6. IEEE, 2016.

[99] S. Rallapalli, A. Ganesan, K. Chintalapudi, V. Padmanabhan, and L. Qiu. Enabling physical analytics in retail stores using smart glasses. In *Proc. of MobiCom'14*.

[100] R. Ravichandran, E. Saba, K.-Y. Chen, M. Goel, S. Gupta, and S. N. Patel. Wibreathe: Estimating respiration rate using wireless signals in natural settings in the home. In *2015 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 131–139. IEEE, 2015.

[101] A. Rubin and J. Ophoff. Investigating adoption factors of wearable technology in health and fitness. In *2018 Open Innovations Conference (OI)*, pages 176–186. IEEE, 2018.

[102] T. Röddiger, D. Wolffram, D. Laubenstein, M. Budde, and M. Beigl. Towards respiration rate monitoring using an in-ear headphone inertial measurement unit. In *Proc. of EarComp'19*, 2019.

[103] S. Schoeppe, S. Alley, W. Van Lippevelde, N. A. Bray, S. L. Williams, M. J. Duncan, and C. Vandelanotte. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, 13(1):127, 2016.

[104] C. Seeger, K. Van Laerhoven, and A. Buchmann. Myhealthassistant: An event-driven middleware for multiple medical applications on a smartphone-mediated body sensor network. *IEEE journal of biomedical and health informatics*, 19(2):752–760, 2015.

[105] R. A. Seguin, C. D. Economos, R. Palombo, R. Hyatt, J. Kuder, and M. E. Nelson. Strength training and older women: a cross-sectional study examining factors related to exercise adherence. *Journal of aging and physical activity*, 18(2):201–218, 2010.

[106] S. Sen, D. Chakraborty, V. Subbaraju, D. Banerjee, A. Misra, N. Banerjee, and S. Mittal. Accommodating user diversity for in-store shopping behavior recognition. In *Proc. ISWC'14*.

[107] S. Sen, A. Misra, V. Subbaraju, K. Grover, M. Radhakrishnan, R. K. Balan, and Y. Lee. I 4 s: capturing shopper's in-store interactions. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers*, pages 156–159. ACM, 2018.

[108] S. Sen, V. Subbaraju, A. MISRA, R. K. Balan, and Y. Lee. The case for smartwatch-based diet monitoring.

[109] L. Shangguan, Z. Zhou, X. Zheng, L. Yang, Y. Liu, and J. Han. Shopminer: Mining customer shopping behavior in physical clothing stores with passive rfids. In *Proc. of Sensys'15*.

[110] L. Shangguan, Z. Zhou, X. Zheng, L. Yang, Y. Liu, and J. Han. Shopminer: Mining customer shopping behavior in physical clothing stores with cots rfid devices. In *Proceedings of the 13th ACM conference on embedded networked sensor systems*, pages 113–125. ACM, 2015.

[111] C. Shen, B.-J. Ho, and M. Srivastava. Milift: Efficient smartwatch-based workout tracking using automatic segmentation. *IEEE Transactions on Mobile Computing*, 17(7):1609–1622, 2018.

[112] S. Sohail. A study of mall shopping behaviour and patronage: Perspectives from an emerging nation. *Available at SSRN 2653640*, 2013.

[113] H. Sorensen, S. Bogomolova, K. Anderson, G. Trinh, A. Sharp, R. Kennedy, B. Page, and M. Wright. Fundamental patterns of in-store shopper behavior. *Journal of Retailing and Consumer Services*, 37:182–194, 2017.

[114] Statista 2017. Global unit sales of headphones and headsets from 2013 to 2017 (in millions). `https://www.statista.com/statistics/327000/worldwide-sales-headphones-headsets/`; [Online Last Accessed: May 2019].

[115] Statista 2019. U.s. fitness center / health club memberships 2000-2017. `https://www.statista.com/statistics/236123/us-fitness-center--health-club-memberships/`; [Online Last Accessed: May 2019].

[116] M. Stiggelbout, M. Hopman-Rock, E. Tak, L. Lechner, and W. van Mechelen. Dropout from exercise programs for seniors: a prospective cohort study. *Journal of Aging and Physical Activity*, 13(4):409–421, 2005.

[117] K. Strohacker, O. Galarraga, and D. M. Williams. The impact of incentives on exercise behavior: a systematic review of randomized controlled trials. *Annals of Behavioral Medicine*, 48(1):92–99, 2013.

[118] V. Subbaraju, S. Sen, A. Misra, S. Chakraborti, and R. Balan. Using infrastructure-provided context filters for efficient fine-grained activity sensing. In *Proc. of PerCom'15*.

[119] M. Sundholm, J. Cheng, B. Zhou, A. Sethi, and P. Lukowicz. Smart-mat: Recognizing and counting gym exercises with low-cost resistive pressure sensing matrix. In *Proc. of ACM UbiComp'14*, pages 373–382. ACM, 2014.

[120] G. J. Székely, M. L. Rizzo, N. K. Bakirov, et al. Measuring and testing dependence by correlation of distances. *The annals of statistics*, 35(6):2769–2794, 2007.

[121] Texas Instruments. Simplelink sensortag from texas instruments, 2018. `http://www.ti.com/ww/en/wireless_connectivity/sensortag/`; [Online Last Accessed: July 2018].

[122] E. Thomaz, I. Essa, and G. D. Abowd. A practical approach for recognizing eating moments with wrist-mounted inertial sensing. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 1029–1040. ACM, 2015.

[123] H. Trinh, Q. Fan, P. Gabbur, and S. Pankanti. Hand tracking by binary quadratic programming and its application to retail activity recognition. In *Proc. of CVPR'12*.

[124] H. Trinh, Q. Fan, J. Pan, P. Gabbur, S. Miyazawa, and S. Pankanti. Detecting human activities in retail surveillance using hierarchical finite state machine. In *Proc. of ICASSP'11*.

[125] S. G. Trost, N. Owen, A. E. Bauman, J. F. Sallis, and W. Brown. Correlates of adults' participation in physical activity: review and update. *Medicine & Science in Sports & Exercise*, 34(12):1996–2001, 2002.

[126] C. C. Tsai, G. Lee, F. Raab, G. J. Norman, T. Sohn, W. G. Griswold, and K. Patrick. Usability and feasibility of pmeb: a mobile phone application for monitoring real time caloric balance. *Mobile networks and applications*, 12(2-3):173–184, 2007.

[127] E. Velloso, A. Bulling, H. Gellersen, W. Ugulino, and H. Fuks. Qualitative activity recognition of weight lifting exercises. In *Proceedings of the 4th Augmented Human International Conference*, pages 116–123. ACM, 2013.

[128] R. H. Venkatnarayan, G. Page, and M. Shahzad. Multi-user gesture recognition using wifi. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, pages 401–413. ACM, 2018.

[129] Vimo Labs Inc. Trackmyfitness, 2018. `http://trackmy.fit`; [Online Last Accessed: December 2018].

[130] T. H. Vu, A. Misra, Q. Roy, K. C. T. Wei, and Y. Lee. Smartwatch-based early gesture detection 8 trajectory tracking for interactive gesture-driven applications. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(1):39, 2018.

[131] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu. E-eyes: device-free location-oriented activity identification using fine-grained wifi signatures. In *Proceedings of the 20th annual international conference on Mobile computing and networking*, pages 617–628. ACM, 2014.

[132] D. E. Whaley and A. F. Schrider. The process of adult exercise adherence: Self-perceptions and competence. *The Sport Psychologist*, 19(2):148–163, 2005.

[133] J. Withall, R. Jago, and K. R. Fox. Why some do but most don't. barriers and enablers to engaging low-income groups in physical activity programmes: a mixed methods study. *BMC public health*, 11(1):507, 2011.

[134] F. Xiao, J. Chen, X. H. Xie, L. Gui, J. L. Sun, and W. none Ruchuan. Seare: A system for exercise activity recognition and quality evaluation based on green sensing. *IEEE Transactions on Emerging Topics in Computing*, 2018.

[135] Z. Yan, V. Subbaraju, D. Chakraborty, A. Misra, and K. Aberer. Energy-efficient continuous activity recognition on mobile phones: An activity-adaptive approach. In *Proc. of ISWC'12*.

[136] C.-W. You, C.-C. Wei, Y.-L. Chen, H.-H. Chu, and M.-S. Chen. Using mobile phones to monitor shopping time at physical stores. *IEEE Pervasive Computing*, (2):37–43, 2011.

[137] G. Zarotis, I. Athanailidis, V. Arvanitidou, and C. Mourtzios. Age-specific reasons for dropping out of the fitness-sport. *Journal of Physical Education and Sport*, 17(2):916, 2017.

[138] Y. Zeng, P. H. Pathak, and P. Mohapatra. Analyzing shopper's behavior through wifi signals. In *Proceedings of the 2nd workshop on Workshop on Physical Analytics (WPA'15)*, pages 13–18. ACM, 2015.

[139] Q. Zhang, D. Wang, R. Zhao, Y. Deng, and Y. Yu. Shopeye: fusing rfid and smartwatch for multi-relation excavation in physical stores. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pages 86–95. ACM, 2019.

[140] X. Zhang, S. Li, R. R. Burke, and A. Leykin. An examination of social influence on shopper behavior using video tracking data. *Journal of Marketing*, 78(5):24–41, 2014.

[141] B. Zhou, M. Sundholm, J. Cheng, H. Cruz, and P. Lukowicz. Never skip leg day: A novel wearable approach to monitoring gym leg exercises. In *Proc. of IEEE PerCom'16*, pages 1–9. IEEE, 2016.